Generalized Cycle Benchmarking Algorithm for Characterizing Midcircuit Measurements

Zhihan Zhang[1,*], Senrui Chen[2,†], Yunchao Liu[3,‡] and Liang Jiang[2,§]

[1] *Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China*

[2] *Pritzker School of Molecular Engineering, The University of Chicago, Chicago, Illinois 60637, USA*

[3] *Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, California 94720, USA*

Midcircuit measurements (MCMs) are crucial ingredients in the development of fault-tolerant quantum computation. While there have been rapid experimental progresses in realizing MCMs, a systematic method for characterizing noisy MCMs is still under exploration. In this work, we develop a cycle benchmarking (CB)-type algorithm to characterize noisy MCMs. The key idea is to use a joint Fourier transform on the classical and quantum registers and then estimate parameters in the Fourier space, analogous to Pauli fidelities used in CB-type algorithms for characterizing the Pauli-noise channel of Clifford gates. Furthermore, we develop a theory of the noise learnability of MCMs, which determines what information can be learned about the noise model (in the presence of state preparation and terminating measurement noise) and what cannot, which shows that all learnable information can be learned using our algorithm. As an application, we show how to use the learned information to test the independence between measurement noise and state-preparation noise in an MCM. Finally, we conduct numerical simulations to illustrate the practical applicability of the algorithm. Similar to other CB-type algorithms, we expect the algorithm to provide a useful toolkit that is of experimental interest.

I. INTRODUCTION

Midcircuit measurements (MCMs) are a central component in quantum error correction and are being developed on various hardware platforms [1–4]. To perform syndrome measurements, one needs to initialize fresh ancilla qubits, apply the syndrome extraction circuit, and measure the ancilla qubits while maintaining the coherence of system qubits. Characterizing the noise model for this process is thus a necessary step for fault-tolerant quantum computation. For quantum gate characterization, a standard approach is to engineer the noise process into Pauli channels [5,6] and various methods have been developed to characterize the Pauli channel [7–12]. However, MCMs are much more complicated than a noisy gate:

*Contact author: zhihan-z21@mails.tsinghua.edu.cn

†Contact author: csenrui@uchicago.edu

‡Contact author: yunchaoliu@berkeley.edu

§Contact author: liang.jiang@uchicago.edu

the measurement readout can be faulty, the postmeasurement state can be faulty, and there can be additional errors on unmeasured qubits. Despite significant recent efforts [13,14], developing general and systematic characterization methods for this more complicated noise model of MCMs still remains open.

In this paper we address this challenge by developing a general algorithm for characterizing MCMs. Our algorithm is a cycle benchmarking (CB)-type protocol [7–10] generalizing CB with interleaved gates [11,12, 15], the state-of-the-art method for characterizing Clifford gates. The key idea behind CB with interleaved gates is the observation that the Fourier transform of the Pauli error rates (known as Pauli fidelities) are eigenvalues of the Pauli channel and therefore are easy to learn via repeated application of the noisy gate. We generalize this view to MCMs by defining Pauli fidelities as a joint Fourier transform on the classical and quantum registers, which provides a means of converting between the physical space (focusing on Pauli errors) and dual space (focusing on Fourier coefficients). A key challenge here is that the Pauli fidelities are no longer eigenvalues of the noise channel and are not directly learnable. We address this by introducing an additional Fourier transform during postprocessing to filter out undesired components. This allows us to use

the CB framework and repeatedly apply MCMs to learn products of Pauli fidelities. See Fig. 2 for a comparison between our algorithm and CB with interleaved gates.

On the other hand, not all parameters of noisy MCMs are learnable in a way that is robust to state preparation and terminating measurement (SPAM [16]) noise, similar to noisy Clifford gates [15], due to the existence of gauge freedom [17]. We give an exact and complete characterization about the learnable parameters of MCMs by generalizing the graph-theoretical framework of Ref. [15], and show that our algorithm is able to learn all learnable parameters of noisy MCMs.

As a concrete application, our algorithm gives a general method to test whether a noisy MCM can be factorized as a measure-and-prepare instrument, which can help understand the underlying physical mechanism for the implementation of MCMs. See Sec. I C for a more detailed summary.

A. General noisy MCMs

We start by defining the noise model associated with MCMs. We consider an ideal MCM as a general Clifford gate G on $n + m$ qubits, followed by a measurement on the n ancilla qubits in the computational basis [see Fig. 1(b)]. The other m qubits are left unmeasured. Note that the quantum register for the n ancilla qubits is still present after the measurement, i.e., we keep the postmeasurement quantum state. We consider this general model because MCMs are often used for qubit reset [4,18–20]. Also note that this setting incorporates subsystem measurements (G is equal to identity), syndrome measurements (G is the syndrome extraction circuit), as well as Clifford gates without measurement ($n = 0$) as special cases.

In the literature of noise characterization, noisy Clifford gates are often modeled as an ideal Clifford gate followed by a Pauli channel (channels where random Pauli error occur) or equivalently, a Pauli channel followed by an ideal Clifford gate. This is due to the randomized compiling techniques [5,6], which can twirl a general CPTP noise channel into a Pauli channel that is more tractable for characterization. Motivated by recent developments in randomized compiling for subsystem measurements [21], we extend these techniques to MCMs. We show that general noisy MCMs can be twirled into the form of an ideal Clifford gate G followed by a uniform stochastic instrument $\mathcal{U}$, which is a highly structured noisy measurement and its physical meaning can be understood as the following three parts:

(1) Suppose we obtain the result k during measurement readout. From this we know that ideally, the n ancilla qubits should be projected into the basis state $|k\rangle$ (throughout the paper, $|k\rangle$ means the computational basis state of index k). However, due to noise in the readout, the ancilla qubits are actually projected into a different basis state $|k + a\rangle$, where a is an n-bit string that came from some probability distribution.

(2) After the ancilla qubits are projected into $|k + a\rangle$, the noise can further affect the postmeasurement state, which becomes $|k + b\rangle$, where b is an n-bit string that came from some probability distribution.

(3) Finally, the measurement process causes an additional random Pauli error P acting on the remaining m qubits due to crosstalk.

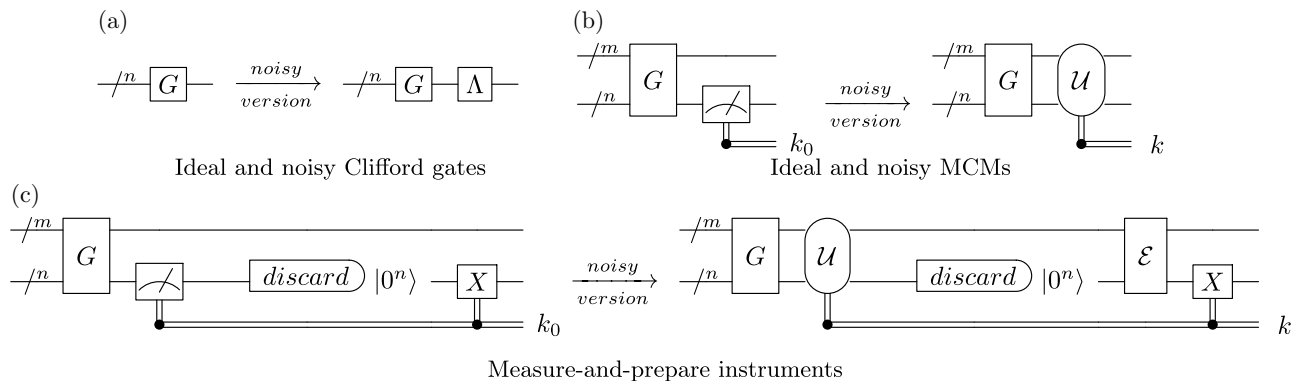The above noise model is parameterized by Pauli error rates, which is a joint probability distribution $p_{a,b}^{P}$

FIG. 1.   Noise models considered in this work. (a) A noisy Clifford gate G is modeled as an ideal gate G followed by a Pauli channel Λ. (b) The noisy version of an MCM, whose corresponding ideal version consists of a Clifford gate G followed by a subsystem measurement, is modeled as the ideal Clifford gate G followed by a uniform stochastic instrument $\mathcal{U}$. The measurement outcome in the noiseless case is denoted k_0 whereas the outcome in the noisy case is denoted k. (c) We also consider a special family of noisy MCMs called measure-and-prepare instruments, in which the noise model can be decomposed into measurement noise and state-preparation noise.

satisfying $\sum_{P\in\mathcal{P}^m, a,b\in\mathbb{Z}_2^n} p_{a,b}^P = 1$ ($\mathcal{P}^m$ is the m-qubit Pauli group).

Our goal is to learn the Pauli error rates, which has $4^{n+m} - 1$ unknown parameters. More specifically, on input density matrix ρ acting on $n + m$ qubits, the output density matrix equals

$$\rho \mapsto \sum_{P\in\mathcal{P}^m, k, a, b\in\mathbb{Z}_2^n} p_{a,b}^P (P \otimes |k + b\rangle \langle k + a|) G\rho G^\dagger$$

$$\times (P \otimes |k + a\rangle \langle k + b|) \otimes |k\rangle \langle k|_R. \quad (1)$$

Here the third register R is a classical register for measurement readout. A key point here is that the Pauli error rates $p_{a,b}^P$ are *independent* from the readout result k; this is a consequence of randomized compiling.

B. Measure-and-prepare instruments

A natural subclass of the general noise model in Eq. (1) is the family of noise channels that can be factorized into measurement noise and state-preparation noise, which we call measure-and-prepare instruments [Fig. 1(c)]. For example, in certain experimental setups, MCMs are realized in a measure-and-prepare manner (e.g., destructive measurements with replaced free ancilla in atom array platforms [4]). After an MCM, the ancillas are initialized in ground state $|0\rangle$, which are then adjusted into the state corresponding to the measurement result through the application of X gates (NOT gates). This sequence of operations results in a special type of MCM and gives us additional structures among the parameters.

A fundamental question about the physical process that implements general MCMs is whether it has the above factorized form. As an application of our result, in Sec. VI we will give a general method for testing whether an unknown MCM can be represented by a measure-and-prepare instrument.

C. Summary of results

In the remainder of the paper, we will first introduce notations and preliminaries in Sec. II. Subsequent sections will detail specific aspects of our results.

1. Fourier transforms and dual-space formalism (Sec. III)

For noisy Clifford gates, standard characterization methods such as some CB-type protocols [7,11,12,15] employ a Fourier transform, converting Pauli error rates into quantities known as Pauli fidelities, which are eigenvalues of the Pauli channel and are easy to learn. Our first technical contribution is to generalize this conversion between the physical space (Pauli errors) and the dual (Fourier) space, by defining a suitable notion of "Pauli fidelities" for

MCMs, which is a joint Fourier transform on the classical and quantum registers.

2. Generalized CB algorithm for characterizing MCMs (Sec. IV)

Similar to CB with interleaved gates, our algorithm learns the Pauli fidelities and reconstructs the Pauli error rates via an inverse Fourier transform. However, the Pauli fidelities of MCMs are no longer eigenvalues of the noise channel, so learning these Pauli fidelities requires new techniques. To address this challenge, we further introduce a key technical idea: an additional Fourier transform on the classical registers during the data-processing procedure, which allows us to learn a single desired component from sums of many components. This completes a generalized version of the CB algorithm to characterize MCMs.

3. Learnability using pattern transfer graph (Sec. V)

Furthermore, we extend the pattern transfer graph technique from Ref. [15] to determine which parameters in the noise model are SPAM-robustly learnable and which are not. We show that our algorithm captures all the information that can be learned, which corresponds to the cycle space of the pattern-transfer graph. Conversely, information that our algorithm cannot learn is inherently unlearnable by any SPAM-robust algorithm. Corresponding numerical simulation results are shown in Fig. 7.

4. Testing measure-and-prepare channels (Sec. VI)

As an application of our algorithm, we show that it can be used to determine whether an arbitrary MCM can be factorized as a measure-and-prepare instrument. This shows that our algorithm can be used to probe the fundamental physical process of the measurement process. Numerical results are shown in Fig. 8.

5. Numerical results (Sec. VII)

Finally, we conduct numerical simulations for the algorithm and applications to support our analysis.

II. NOTATIONS AND PRELIMINARIES

A. Pauli operators, Clifford gates

In this paper, $\mathcal{P}^m := \{I, X, Y, Z\}^m$ represents the m-qubit Pauli group (modulus phase), $\mathcal{C}$ represents the group of single-qubit Clifford gates.

B. Pauli operator with vector exponents

For $P \in \mathcal{P}^1$ and $x \in \mathbb{Z}_2^n$, we use P^x to represent the n-qubit Pauli operator that act as I on the ith qubit if the ith bit of x is 0, and act as P on the ith qubit if the ith bit of x is 1.

C. Inner product of Pauli operators

For $P, Q \in \mathcal{P}^m$, we define inner product $\langle P, Q \rangle$ to be 0 if Pauli operators P and Q commute, and 1 otherwise [22]. That is, $PQ = (-1)^{\langle P, Q \rangle} QP$.

D. Pauli weight patterns

Following Ref. [15], for $P \in \mathcal{P}^m$, we define its Pauli weight pattern pt(P) to be the m-bit string that is 0 on the ith bit if P act as identity on the ith qubit, and is 1 on the ith bit if P act as nonidentity on the ith qubit.

E. Quantum channels

Calligraphic letters represent channels, for example, $\mathcal{E}, \mathcal{M}, \mathcal{T}, \mathcal{U}$. Specifically, $\mathcal{G}, \mathcal{H}, \mathcal{P}, \mathcal{X}, \mathcal{Z}$ represent specific channels of applying gates of the corresponding normal letters G, H, P, X, Z. For example, $\mathcal{G}(\rho) = G \rho G^\dagger$.

F. Pauli channels

Pauli channels are represented by Λ exclusively. They are channels of applying a random Pauli operator according to a probability distribution,

$$\Lambda = \sum_{P \in \mathcal{P}^m} p^P \mathcal{P}, \qquad (2)$$

where $\{p^P\}_{P \in \mathcal{P}^m}$ are called **Pauli error rates**.

G. Super-operator formalism

General treatment of the super-operator formalism can be found in, e.g., Ref. [23]. In short, vectorization refers to the isomorphism $|i\rangle \langle j| \mapsto |ij\rangle$. The inner product over the vectorized space is then given by $\langle\!\langle O | P \rangle\!\rangle = \text{Tr}(O^\dagger P)$. In this paper, for $k \in \mathbb{Z}_2^n$, $|k\rangle\!\rangle$ is the vectorization of computational basis state $|k\rangle \langle k|$. And for $P \in \mathcal{P}^m$, double brackets $|P\rangle\!\rangle$ represents the vectorization of Pauli operator P.

H. Pauli fidelities

Recall from Eq. (2) that a Pauli channel is of the form $\Lambda = \sum_{P \in \mathcal{P}^m} p^P \mathcal{P}$. Note that the Pauli operators are eigenvectors of the channel:

$$\Lambda |Q\rangle\!\rangle = \sum_{P \in \mathcal{P}^m} p^P \mathcal{P} |Q\rangle\!\rangle = \sum_{P \in \mathcal{P}^m} (-1)^{\langle P, Q \rangle} p^P |Q\rangle\!\rangle. \quad (3)$$

The corresponding eigenvalues are defined as Pauli fidelities [7]

$$\lambda^Q = \sum_{P \in \mathcal{P}^m} (-1)^{\langle P, Q \rangle} p^P. \qquad (4)$$

For normalized Pauli channels, the trace-preserving condition $\sum_{p \in \mathcal{P}^m} p^P = 1$ translates into $\lambda^I = 1$. The inverse

transformation is given by

$$p^P = \frac{1}{4^m} \sum_{Q \in \mathcal{P}^m} (-1)^{\langle P, Q \rangle} \lambda^Q. \qquad (5)$$

Using the Pauli fidelities, we can write the Pauli channel in the dual space

$$\Lambda = \frac{1}{2^m} \sum_{Q \in \mathcal{P}^m} \lambda^Q |Q\rangle\!\rangle \langle\!\langle Q|. \qquad (6)$$

This uses the Pauli-transfer-matrix representation, which is diagonal for Pauli channels. We say that the Pauli error rates lie in the physical space because they represent physical errors. Correspondingly, we say that the Pauli fidelities lie in the dual space.

I. Quantum instruments

Mathematically, general MCMs are modeled as quantum instruments, which are sets of completely positive, trace nonincreasing maps $\{\mathcal{E}_j\}$ such that $\sum_j \mathcal{E}_j$ is trace preserving. The action of it on a density operator ρ is [24]

$$\rho \mapsto \sum_j \mathcal{E}_j(\rho) \otimes |j\rangle \langle j|. \qquad (7)$$

That is, with probability $p_j = \text{Tr}(\mathcal{E}_j(\rho))$, outcome j is observed and the output state becomes $(1/p_j)\mathcal{E}_j(\rho)$.

III. NOISE MODELS AND THEIR FOURIER TRANSFORM

A. Uniform stochastic instrument

As mentioned in the introduction, the motivation for modeling MCMs as uniform stochastic instruments is randomized compiling. But we will leave the details to Appendix A 1 and begin our discussion with a formal definition of the uniform stochastic instruments.

Using the super-operator formalism, we can rewrite our noise model in Eq. (1) to the form of quantum instruments.

Definition 1 (Uniform stochastic instruments, Eq. (17) of Ref. [21]). In physical space, the uniform stochastic instrument is the set $\{\mathcal{U}_k\}$ where each $\mathcal{U}_k$ is expressed by the following trace nonincreasing map

$$\mathcal{U}_k := \sum_{a,b \in \mathbb{Z}_2^n, P \in \mathcal{P}^m} p_{a,b}^P \mathcal{P} \otimes |k+b\rangle\!\rangle \langle\!\langle k+a| \qquad (8)$$

$$= \sum_{a,b \in \mathbb{Z}_2^n} \Lambda_{a,b} \otimes |k+b\rangle\!\rangle \langle\!\langle k+a| \qquad (9)$$

here $p_{a,b}^P$ is a joint probability distribution on $\mathbb{Z}_2^n \times \mathbb{Z}_2^n \times \mathcal{P}^m$ called Pauli error rates and $\Lambda_{a,b}$ are unnormalized Pauli

channels indexed by a, b (note that they are independent of k).

The overall noise model in Eq. (1) can then be re-expressed as

$$|\rho\rangle\!\rangle \mapsto \sum_{k \in \mathbb{Z}_2^n} \mathcal{U}_k \mathcal{G} |\rho\rangle\!\rangle \otimes |k\rangle\!\rangle_R. \tag{10}$$

Let $\mathcal{T}_k := \mathcal{U}_k \mathcal{G}$. As the Pauli channel $\Lambda_{a,b}$ is independent of k, in the following we will often drop the third register $|k\rangle\!\rangle_R$ and work with a specific map $\mathcal{T}_k$. Then we can see that $\{\mathcal{U}_k\}$ and $\{\mathcal{T}_k\}$ are quantum instruments since $\sum_{a,b \in \mathbb{Z}_2^n, P \in \mathcal{P}^m} p_{a,b}^P = 1$.

Now we generalize the physical vs dual space picture to uniform stochastic instruments. We define the Pauli fidelities $\lambda_{x,y}^Q$ as the Fourier transform of the Pauli error rates

$$\lambda_{x,y}^Q = \sum_{a,b \in \mathbb{Z}_2^n, P \in \mathcal{P}^m} (-1)^{a \cdot x + b \cdot y + \langle P, Q \rangle} p_{a,b}^P. \tag{11}$$

We note that the Pauli fidelities defined above are different from the Pauli fidelities of the unnormalized Pauli channels in Eq. (9) as there is an additional Fourier transformation

on the subscript a and b. The role of the Pauli fidelities will become clear after the next lemma.

In fact, our protocol will first learn the Pauli fidelities and then use the inverse transformation

$$p_{a,b}^P = \frac{1}{4^{n+m}} \sum_{x,y \in \mathbb{Z}_2^n, Q \in \mathcal{P}^m} (-1)^{a \cdot x + b \cdot y + \langle P, Q \rangle} \lambda_{x,y}^Q \tag{12}$$

to obtain the Pauli error rates.

Using Pauli fidelities, we can write the uniform stochastic instrument and thus the instrument $\{\mathcal{T}_k\}$ of our model in the dual space.

Lemma 1. Let $\mathcal{U}_k$ be a uniform stochastic instrument defined in Eq. (9), and let $\mathcal{T}_k = \mathcal{U}_k \mathcal{G}$ where $\mathcal{G}$ is a (Clifford) unitary channel. Then $\mathcal{T}_k$ can be expressed in dual space as

$$\mathcal{T}_k = \frac{1}{2^{2n+m}} \sum_{x,y \in \mathbb{Z}_2^n, Q \in \mathcal{P}^m} (-1)^{k \cdot (x+y)} \lambda_{x,y}^Q |Q \otimes Z^y\rangle\!\rangle$$
$$\times \langle\!\langle \mathcal{G}^\dagger (Q \otimes Z^x)|. \tag{13}$$

Proof. For a uniform stochastic instrument, we have

$$\mathcal{U}_k = \sum_{a,b \in \mathbb{Z}_2^n} \Lambda_{a,b} \otimes |k+b\rangle\!\rangle \langle\!\langle k+a| \tag{14}$$

$$= \sum_{a,b \in \mathbb{Z}_2^n} \Lambda_{a,b} \otimes \left(\frac{1}{2^n} \sum_{Q \in \mathcal{P}^n} |Q\rangle\!\rangle \langle\!\langle Q| \right) |k+b\rangle\!\rangle \langle\!\langle k+a| \left(\frac{1}{2^n} \sum_{P \in \mathcal{P}^n} |P\rangle\!\rangle \langle\!\langle P| \right) \tag{15}$$

$$= \sum_{a,b \in \mathbb{Z}_2^n} \Lambda_{a,b} \otimes \left(\frac{1}{2^n} \sum_{y \in \mathbb{Z}_2^n} |Z^y\rangle\!\rangle \langle\!\langle Z^y| \right) |k+b\rangle\!\rangle \langle\!\langle k+a| \left(\frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} |Z^x\rangle\!\rangle \langle\!\langle Z^x| \right) \tag{16}$$

$$= \frac{1}{2^{2n}} \sum_{x,y \in \mathbb{Z}_2^n} (-1)^{k \cdot (x+y)} \sum_{a,b \in \mathbb{Z}_2^n} (-1)^{a \cdot x + b \cdot y} \Lambda_{a,b} \otimes |Z^y\rangle\!\rangle \langle\!\langle Z^x| \tag{17}$$

$$= \frac{1}{2^{2n}} \sum_{x,y \in \mathbb{Z}_2^n} (-1)^{k \cdot (x+y)} \sum_{a,b \in \mathbb{Z}_2^n} (-1)^{a \cdot x + b \cdot y} \left(\frac{1}{2^m} \sum_{P,Q \in \mathcal{P}^m} (-1)^{\langle P, Q \rangle} p_{a,b}^P |Q\rangle\!\rangle \langle\!\langle Q| \right) \otimes |Z^y\rangle\!\rangle \langle\!\langle Z^x| \tag{18}$$

$$= \frac{1}{2^{2n+m}} \sum_{x,y \in \mathbb{Z}_2^n, Q \in \mathcal{P}^m} (-1)^{k \cdot (x+y)} \left(\sum_{a,b \in \mathbb{Z}_2^n, P \in \mathcal{P}^m} (-1)^{a \cdot x + b \cdot y + \langle P, Q \rangle} p_{a,b}^P \right) |Q \otimes Z^y\rangle\!\rangle \langle\!\langle Q \otimes Z^x| \tag{19}$$

$$= \frac{1}{2^{2n+m}} \sum_{x,y \in \mathbb{Z}_2^n, Q \in \mathcal{P}^m} (-1)^{k \cdot (x+y)} \lambda_{x,y}^Q |Q \otimes Z^y\rangle\!\rangle \langle\!\langle Q \otimes Z^x|. \tag{20}$$

Equation (15) is a basis change. Equation (16) uses the fact that Pauli operators that have nonzero inner product (over the vectorized space) with a computational basis state must be a tensor product of I and Zs. The phases in Eq. (17) comes from the inner products $\langle\langle Z^y \mid k+b \rangle\rangle$ and $\langle\langle k+a \mid Z^x \rangle\rangle$. Equation (18) rewrites the Pauli channels using Eq. (6).

As can be seen from Eq. (20), the Pauli fidelities are *not* eigenvalues of the uniform stochastic instrument. They connect the Pauli operators $Q \otimes Z^y$ and $Q \otimes Z^x$.

Now put an ideal gate G before this instrument. Since $\mathcal{G}^\dagger(Q \otimes Z^x)$ becomes $Q \otimes Z^x$ after applying $\mathcal{G}$, $\langle\langle Q \otimes Z^x | \mathcal{G} = \langle\langle \mathcal{G}^\dagger(Q \otimes Z^x)|$, we get the result. ∎

When noise rates are small, we would expect that $p_{0,0}^I$ is close to 1, while other Pauli error rates are small. This means that all Pauli fidelities will be close to 1. Note that the trace preserving condition $\sum_{P \in \mathcal{P}^m, a,b \in \mathbb{Z}_2^n} p_{a,b}^P = 1$ translates into $\lambda_{0,0}^I = 1$ here.

TABLE I. Summary comparing physical space and dual space, integrating our findings with previous works. Pointers are provided for relevant definitions and details. The concept of cycle space and cut space will be discussed in Sec. IV B.

		Physical space (Pauli error rate)	Dual space (Pauli fidelity)
Noisy Clifford gates [Fig. 1(a)]	Pauli channels (with Clifford gates)	$\sum_{P \in \mathcal{P}^m} p^P \mathcal{P}\mathcal{G}$ [Eq. (2)]	$\frac{1}{2^m} \sum_{Q \in \mathcal{P}^m} \lambda^Q |Q\rangle\rangle \langle\langle \mathcal{G}^\dagger(Q)|$ [Eq. (21)]
	learnability	same as the Pauli fidelities (see Sec. D)	cycle space are learnable, cut space are not (Theorem 2 in [15])
	conversion formula	$p^P = \frac{1}{4^m} \sum_{Q \in \mathcal{P}^m} (-1)^{\langle P,Q \rangle} \lambda^Q$ [Eq. (5)]	$\lambda^Q = \sum_{P \in \mathcal{P}^m} (-1)^{\langle P,Q \rangle} p^P$ [Eq. (4)]
Noisy MCMs [Fig. 1(b)]	uniform stochastic instruments (with Clifford gates)	$\sum_{\substack{a,b \in \mathbb{Z}_2^n \\ P \in \mathcal{P}^{\tilde{m}}}} p_{a,b}^P (\mathcal{P} \otimes |k+b\rangle\rangle \langle\langle k+a|)\mathcal{G}$ [Eq. (8)]	$\frac{1}{2^{2n+m}} \sum_{\substack{x,y \in \mathbb{Z}_2^n \\ Q \in \mathcal{P}^{\tilde{m}}}} (-1)^{k \cdot (x+y)} \lambda_{x,y}^Q |Q \otimes Z^y\rangle\rangle \langle\langle \mathcal{G}^\dagger(Q \otimes Z^x)|$ [Eq. (13)]
	learnability	different from the Pauli fidelities, partially characterized (see Sec. V B)	cycle space are learnable, cut space are not (see Sec. V A)
	conversion formula	$p_{a,b}^P = \frac{1}{4^{n+m}} \sum_{\substack{x,y \in \mathbb{Z}_2^n \\ Q \in \mathcal{P}^{\tilde{m}}}} (-1)^{a \cdot x + b \cdot y + \langle P,Q \rangle} \lambda_{x,y}^Q$ [Eq. (12)]	$\lambda_{x,y}^Q = \sum_{\substack{a,b \in \mathbb{Z}_2^n \\ P \in \mathcal{P}^{\tilde{m}}}} (-1)^{a \cdot x + b \cdot y + \langle P,Q \rangle} p_{a,b}^P$ [Eq. (11)]
Noisy measure-and-prepare [Fig. 1(c)]	measure and prepare instruments (with Clifford gates)	$\sum_{\substack{a,b \in \mathbb{Z}_2^n \\ P \in \mathcal{P}^{\tilde{m}} \\ P_1 \cdot P_2 = P}} q_a^{P_1} r_b^{P_2} (\mathcal{P} \otimes |k+b\rangle\rangle \langle\langle k+a|)\mathcal{G}$ [Eq. (B15)]	$\frac{1}{2^{2n+m}} \sum_{\substack{x,y \in \mathbb{Z}_2^n \\ Q \in \mathcal{P}^{\tilde{m}}}} (-1)^{k \cdot (x+y)} \zeta_x^Q \xi_y^Q |Q \otimes Z^y\rangle\rangle \langle\langle \mathcal{G}^\dagger(Q \otimes Z^x)|$ [Eq. (B5)]
	learnability	same as uniform stochastic instruments, but overly constrained (see Sec. VI)	same as uniform stochastic instruments, but overly constrained (see Sec. VI)
	conversion formula	$q_a^P = \frac{1}{2^{n+2m}} \sum_{\substack{x \in \mathbb{Z}_2^n \\ Q \in \mathcal{P}^{\tilde{m}}}} (-1)^{a \cdot x + \langle P,Q \rangle} \zeta_x^Q$ $r_b^P = \frac{1}{2^{n+2m}} \sum_{\substack{x \in \mathbb{Z}_2^n \\ Q \in \mathcal{P}^{\tilde{m}}}} (-1)^{b \cdot x + \langle P,Q \rangle} \xi^Q$ [Eq. (B7)]	$\zeta_x^Q = \sum_{\substack{a \in \mathbb{Z}_2^n \\ P \in \mathcal{P}^{\tilde{m}}}} (-1)^{a \cdot x + \langle P,Q \rangle} q_a^P$ $\xi_x^Q = \sum_{\substack{a \in \mathbb{Z}_2^n \\ P \in \mathcal{P}^{\tilde{m}}}} (-1)^{b \cdot x + \langle P,Q \rangle} r_b^P$ [Eq. (B8)]

B. Measure-and-prepare instruments

It turns out that the additional structure for measure-and-prepare instruments over uniform stochastic instruments is that the Pauli error rates p factorizes according to $p_{a,b}^P = \sum_{P_1 \cdot P_2 = P} q_a^{P_1} r_b^{P_2}$, where q and r are two probability distributions and the Pauli fidelities factorizes according to $\lambda_{x,y}'^Q = \zeta_x^Q \xi_y^Q$. We defer the proofs to Sec. B.

A comparison between the physical and dual space is listed in Table I, which summarizes our results.

IV. ALGORITHMS FOR NOISE CHARACTERIZATION

A. Generalizing CB-type protocols

For intuition, we will start with the special case of $n = 0$ in our CB-type protocol. In this scenario, our protocol simplifies to the CB with interleaved gates protocol for benchmarking noisy Clifford gates. Variants of this protocol have been discussed in previous works, such as Refs. [7,10,12], with the exact version employed in Ref. [15] and utilized in Sec. SV of Ref. [11]. However, given that this is a special case, we will not dig into a comprehensive treatment here.

In CB with interleaved gates, a noisy Clifford gate to be characterized is modeled as an ideal Clifford gate followed by a Pauli noise channel. Naturally we would want to learn the Pauli error rates of the error. However, concatenation of Pauli channels corresponds to convolution of Pauli error rates, making it complicated to work with. Therefore, a Fourier transformation is employed and the theory evolves within the dual space. In the dual space, combine Eq. (6) to a Clifford gate $\mathcal{G}$ we get the noisy version of it,

$$\widetilde{\mathcal{G}} = \Lambda \circ \mathcal{G} = \frac{1}{2^m} \sum_{Q \in \mathcal{P}^m} \lambda^Q |Q\rangle\!\rangle \langle\!\langle \mathcal{G}^\dagger(Q)|. \quad (21)$$

That is, the $\mathcal{G}^\dagger(Q)$ component of the input density matrix is converted to the Q component of the output density matrix and shrunk by a factor of λ^Q at the same time. CB with interleaved gates repeatedly applies such noisy gates for multiple times and measures the component in the final density matrix. In addition, single-qubit Clifford gates are interleaved between the midcircuit measurements to ensure the correct concatenation of the transitions. Take the simplest case where there are no interleaving single-qubit Clifford gates as an example. Start with a noisy input state $(I + \lambda_S^P P)/2^m$ (λ_S^P is an unknown state-preparation noise parameter), repeatedly applying the noisy Clifford gate l times gives

$$\widetilde{\mathcal{G}}^l \left(\frac{I + \lambda_S^P P}{2^m} \right) = \frac{I + \lambda_S^P \lambda^{\mathcal{G}(P)} \cdots \lambda^{\mathcal{G}^l(P)} \mathcal{G}^l(P)}{2^m}. \quad (22)$$

If we then measure (with noise) the observable $\mathcal{G}^l(P)$, the expected value we get will be $\lambda_S^P \lambda^{\mathcal{G}(P)} \cdots \lambda^{\mathcal{G}^l(P)} \lambda_M^{\mathcal{G}^l(P)}$, where $\lambda_M^{\mathcal{G}^l(P)}$ is an unknown measurement noise. Suppose P is invariant under $\mathcal{G}$, the expectation value becomes $\lambda_S^P (\lambda^P)^l \lambda_M^P$. In this case, by conducting experiments with different l and conduct an exponential fitting, one can estimate λ^P independently of the SPAM noise. For generic P that is not invariant under $\mathcal{G}$, since $\{\mathcal{G}(P), \mathcal{G}^2(P), \ldots\}$ is periodic, similar approaches can be used to estimate certain product of Pauli fidelities SPAM robustly.

Inspired by this, we interleave single-qubit Clifford gates between noisy MCMs. An illustrative example circuit is presented in Fig. 2. However, in our MCM scenario, Pauli operators cease to be eigenvectors (cf. Lemma 1 and note the summation over y therein). In fact, prior to the characterization, the eigenvectors are unknown, so we cannot prepare and measure the eigenvectors as in the CB with interleaved gates case. Hence we still choose to prepare and measure Pauli observables, but now simple interleaving is insufficient because multiple

(a)

1. Prepare Pauli eigenstate
2. Apply repeated Clifford gates
3. Estimate Pauli observable
4. Obtain sum of log Pauli fidelities
5. Perform Fourier transform, obtain Pauli error

CB with interleaved gates

(b)

1. Prepare Pauli eigenstate
2. Apply repeated MCMs
3. Estimate Pauli observable
4. *Perform Fourier transform on measurement outcome*
5. Obtain sum of log Pauli fidelities
6. Perform Fourier transform, obtain Pauli error
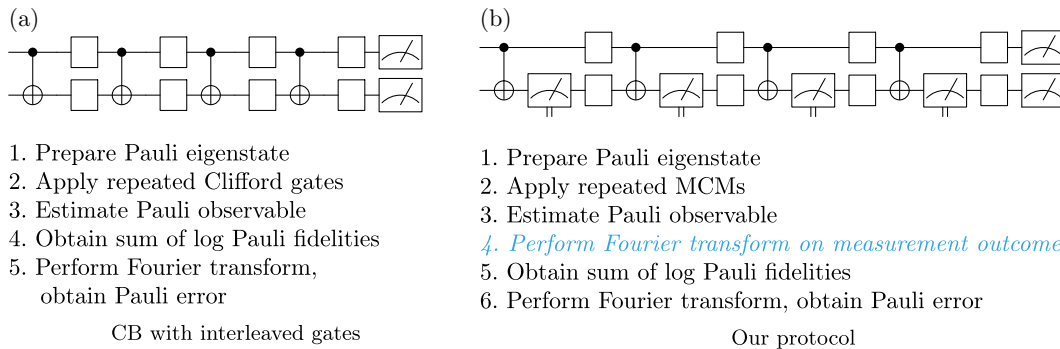
Our protocol

FIG. 2.   A comparison between CB with interleaved gates (a) and our protocol for characterizing MCMs (b). Here, blank gates represent single-qubit Clifford gates. The quantum circuit in (a) is used to characterize a noisy CNOT gate, while the circuit in (b) is used to characterize a noisy MCM (measuring the target qubit after applying a CNOT). The key difference between the two algorithms is an additional Fourier transform on measurement outcome (highlighted in blue).
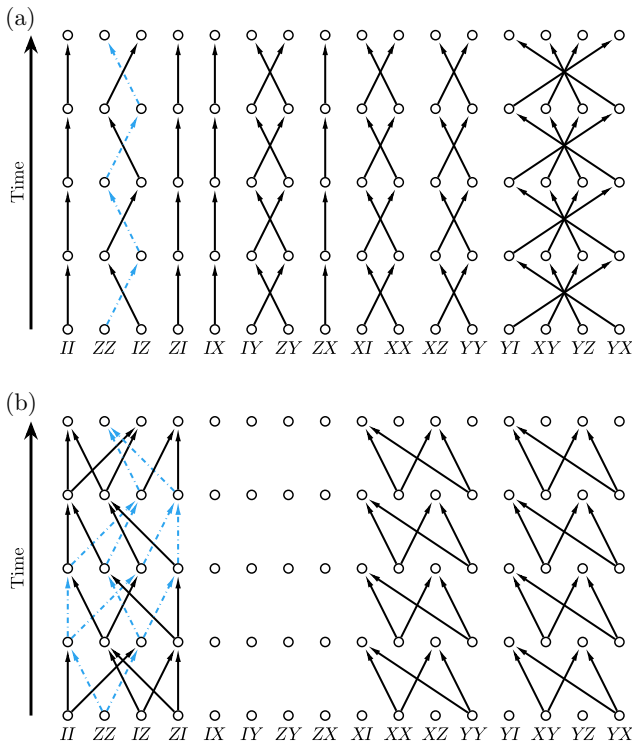
FIG. 3. Comparison between the composition of noisy Clifford gates [$\widetilde{\mathcal{G}}$ in Eq. (21)] and noisy MCMs [$\mathcal{T}$ in Eq. (13)] in the dual space. (a) Composition of noisy Clifford gates ($G = $ CNOT, $m = 2, n = 0$, with no interleaving single-qubit Clifford gates) in the dual space. The noisy Clifford gates take one Pauli operator to one Pauli operator, so different paths cannot have the same start point and end point in the same time (highlighted by blue dash-dot lines), and thus they do not superpose with each other. (b) Composition of noisy MCMs ($G = $ CNOT, $m = n = 1$, with no interleaving single-qubit Clifford gates) in the dual space. Multiple transitions are happening in the same time, so fix a start point and an end point, there may be many paths (highlighted by blue dash-dot lines) between them and these paths will superpose together.

transitions occur simultaneously, superposing together and yielding complicated results. In detail, if we perform many experiments, group the terminating measurement results according to the MCM results and average them, the corresponding conditional expectation values are compromised of many terms, each term is a product of Pauli fidelities. This new phenomenon is illustrated in Fig. 3.

Details of the figure may be understood more easily by referencing Sec. IV C. Each column of nodes corresponds to a vectorization of a Pauli operator. An arrow from column P to column Q represents $|Q\rangle\!\rangle\langle\!\langle P|$ together with the associated coefficients, which are Pauli fidelities up to signs. In Fig. 3(a), arrows start at $\mathcal{G}^\dagger(Q)$ and end at Q for $Q \in \mathcal{P}^m$. Take $Q = IZ$ for example, we have $\mathcal{G}^\dagger(Q) = $ CNOT$^\dagger \cdot IZ \cdot$ CNOT $= ZZ$ and $Q = IZ$, so we get arrows from ZZ to IZ. Similarly, in Fig. 3(b), arrows

start at $\mathcal{G}^\dagger(Q \otimes Z^x)$ and end at $Q \otimes Z^y$ for $x, y \in \mathbb{Z}_2^n$ and $Q \in \mathcal{P}^m$. Take $Q = I$, $x = 1$, $y \in \{0, 1\}$, for example, we have $\mathcal{G}^\dagger(Q \otimes Z^x) = $ CNOT$^\dagger \cdot IZ \cdot$ CNOT $= ZZ$, $Q \otimes Z^y \in \{II, IZ\}$, so we get arrows from ZZ to II and IZ. Arrows of a specific row represent all elements of $\widetilde{\mathcal{G}}$ or $\mathcal{T}$ of a specific time. By distributive law, the composition (the product) contains all possible combinations of edges from the channels. However, since the vectorized Pauli operators are pairwise orthogonal, only consecutive edges (paths) may have nonzero contribution to the product since they correspond to the product of the form $|R\rangle\!\rangle\langle\!\langle Q|Q\rangle\!\rangle\langle\!\langle P|\cdots$, with coefficients omitted. Some of the edges cannot form consecutive paths, indicating that their contribution is completely eliminated by subsequent randomized compiling, so if we want to learn them we should change the interleaving single-qubit Clifford gates to make them consecutive. If we specify a Pauli component of the input state and a terminating measurement, only paths of the specific starting point and end point will have contribution to the probabilities.

To disentangle this superposition of different transition paths so as to extract a single term, we need to apply another Fourier transformation on the measurement results in our data-processing procedure. It involves aggregating the averages while incorporating the Fourier coefficients. This is one of our new ideas, and further details will be provided in Sec. IV C.

B. Pattern-transfer graph

Before formally introducing our protocol, we first need to define the *pattern-transfer graph*, which is a directed graph (V, E) where vertices $V = \mathbb{Z}_2^{n+m}$ corresponds to Pauli weight patterns and edges $E = \{e_{x,y}^Q := (\text{pt}(\mathcal{G}^\dagger(Q \otimes Z^x)), \text{pt}(Q \otimes Z^y))|x, y \in \mathbb{Z}_2^n, Q \in \mathcal{P}^m\}$ corresponds to log Pauli fidelities. An example of the pattern transfer for $G = $ CNOT ($m = n = 1$) is shown in Fig. 4. Our graph can be viewed as a generalization of the pattern-transfer graph defined in Ref. [15]. For comparison, the pattern-transfer graph for $G = $ CNOT in the CB with interleaved gates' case ($m = 2, n = 0$) is also included [25].

With pattern-transfer graph in hand, we need some tools from graph theory [26,27] to exploit its power. A (directed) path is an alternating sequence of vertices and edges, $v_0, e_1, v_1, \ldots, e_n, v_n$ such that $e_i = (v_{i-1}, v_i)$.

Let $\epsilon_i \in \mathbb{R}$. A *0 chain* is a formal linear combination of vertices $\sum \epsilon_i v_i$, while a *1 chain* is a formal linear combination of edges $\sum \epsilon_i e_i$. For example, $- \textcircled{01} + 2\,\textcircled{10} - \textcircled{11}$ is a 0 chain and $e_{1,1}^Z + 2e_{0,0}^X$ is a 1 chain (in this section, all graph theory examples are considered for Fig. 4 left). We emphasize that for pattern-transfer graph, vertices are elements in $\mathbb{Z}_2^{n+m}$ (Pauli weight patterns), but the readers shall not confuse 0 chain with addition in $\mathbb{Z}_2^{n+m}$ as we never perform addition on Pauli weight patterns in $\mathbb{Z}_2^{n+m}$ sense. In the
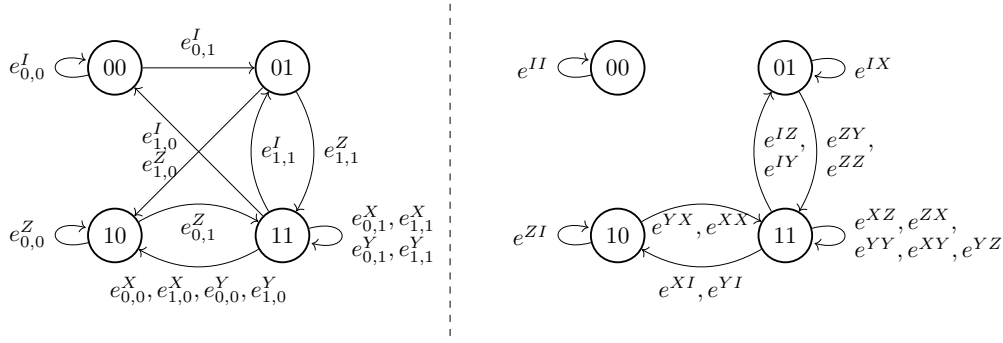
FIG. 4. Left: pattern transfer graph for $G = $ CNOT. Here $m = n = 1$. The ancilla is the target qubit. This is the case for X measurements using ancilla. As an example, the start point for $e^I_{0,1}$ is $\mathrm{pt}(\mathcal{G}^\dagger(II)) = 00$, and the end point for $e^I_{0,1}$ is $\mathrm{pt}(IZ) = 01$. Right: pattern transfer graph for $G = $ CNOT. Here $m = 2$, $n = 0$. No qubits are measured, so the situation degenerates into the CB with interleaved gates case. This graph can be found in [15]. Note that the rules are slightly different so the labels are different. See the footnote for explanation.

previous example of 0 chain, we have circled the vertices for the distinction. The *edge space* C is the vector space over $\mathbb{R}$ formed by 1 chains, together with inner product $\langle \sum \epsilon_i e_i, \sum \epsilon'_i e_i \rangle = \sum \epsilon_i \epsilon'_i$. For example, $e^Z_{1,1} + 2e^X_{0,0} \in C$ and $\langle e^Z_{1,1} + 2e^X_{0,0}, e^I_{1,1} + e^Z_{1,1} \rangle = 1$.

The *boundary operator* ∂ is a linear operator that sends 1 chains to 0 chains such that if $e = (u, v)$, then $\partial e = v - u$. For example, $\partial(e^Z_{1,1} + 2e^X_{0,0}) = (\textcircled{11} - \textcircled{01}) + 2(\textcircled{10} - \textcircled{11}) = -\textcircled{01} + 2\textcircled{10} - \textcircled{11}$. The *coboundary operator* δ is a linear operator that sends 0 chains to 1 chains such that $\delta(v) = \sum \epsilon_i e_i$ where $\epsilon_i = 1$ if $e_i = (u, v)$ for some $u \neq v$, $\epsilon_i = -1$ if $e_i = (v, u)$ for some $u \neq v$, and $\epsilon_i = 0$ otherwise. For example, $\delta(\textcircled{00}) = e^I_{1,0} - e^I_{0,1}$. A *cycle vector* is a 1 chain with boundary 0. The *cycle space* Z is the subspace of C formed by all cycle vectors. For example, $\partial(e^I_{1,1} + e^Z_{1,1}) = (\textcircled{01} - \textcircled{11}) + (\textcircled{11} - \textcircled{01}) = 0$, so $e^I_{1,1} + e^Z_{1,1} \in Z$. A *cut vector* is a coboundary of some 0 chain. The *cut space* U is the subspace of C formed by all cut vectors. For example, $\delta(\textcircled{00}) = e^I_{1,0} - e^I_{0,1} \in U$. The following lemma describes the relation between C, Z, and U.

Lemma 2 ([27], Sec. II.3, Theorem 9). The edge space C is the orthogonal direct sum of cycle space Z and cut space U.

C. Protocol details

Now we present our protocol in details. For this we make use of the pattern-transfer graph. For any given path $v_0, e^{Q_1}_{x_1,y_1}, v_1, \ldots, e^{Q_l}_{x_l,y_l}, v_l$, since $\mathrm{pt}(Q_i \otimes Z^{y_i}) = v_i = \mathrm{pt}(\mathcal{G}^\dagger(Q_{i+1} \otimes Z^{x_{i+1}}))$, there exists $H_i \in \mathcal{C}^{\otimes n+m}$ (recall that $\mathcal{C}$ is the group of single-qubit Clifford gates, so H_is are tensor products of single-qubit Clifford gates) such that

$$\mathcal{H}_i(Q_i \otimes Z^{y_i}) = \mathcal{G}^\dagger(Q_{i+1} \otimes Z^{x_{i+1}}). \qquad (23)$$

We claim that the following protocol is able to estimate

$$\log\left(\lambda^{v_l}_M \langle\!\langle \mathcal{G}^\dagger(Q_1 \otimes Z^{x_1})|\rho\rangle\!\rangle \prod_{i=1}^l \lambda^{Q_i}_{x_i,y_j}\right)$$

$$\equiv \log\langle\!\langle \mathcal{G}^\dagger(Q_1 \otimes Z^{x_1})|\rho\rangle\!\rangle + \sum_{i=1}^l \log\lambda^{Q_i}_{x_i,y_i} + \log\lambda^{v_l}_M. \qquad (24)$$

(1) Prepare Pauli eigenstate: prepare an arbitrary but fixed state ρ. Preferably, it has a large overlap with the Pauli eigenstate $(I + Q_1 \otimes Z^{x_1})/2^{n+m}$.

(2) Apply repeated MCMs: for $i = 1, \ldots, l$, perform a compiled MCM (we need to apply randomized compiling on the MCM to ensure that it follows our noise model, see Appendix A 1 for details), record the result m_i, and then apply gate H_i (no need to apply gate H_i for $i = l$).

(3) Estimate Pauli observable: perform a compiled terminating measurement (again we need to apply randomized compiling on the terminating measurement to ensure that it follows our noise model, see Appendix A 2 for details) for observable $Q_l \otimes Z^{y_l}$ and record the result r.

(4) Perform Fourier transform on measurement outcome: repeat the above procedure multiple times and estimate the expectation value $s = \mathbb{E}\left[(-1)^{\sum_{i=1}^l m_i \cdot (x_i+y_i)} r\right]$.

(5) Obtain sum of log Pauli fidelities: output $\log s$.

Figure 5 illustrates the circuit used in the protocol. It turns out to be mathematically convenient if we modify the protocol a little bit by adding another experiment to cancel out the state-preparation noise. The modified protocol can
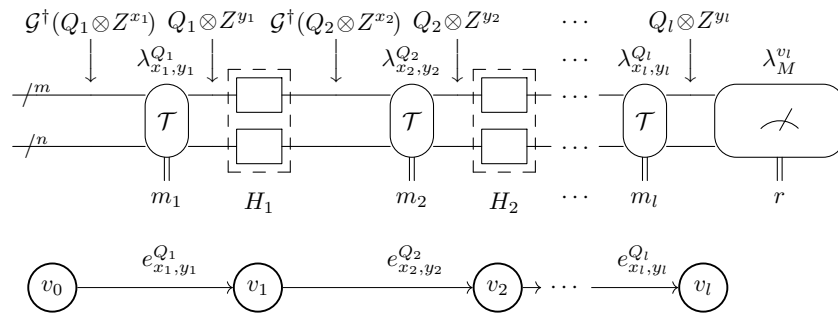
FIG. 5. Main circuit used in the protocol and the corresponding "walk" on the pattern transfer graph. We interleave MCMs with the precomputed H_is and perform a terminating measurement in the end. The $\mathcal{G}^\dagger(Q_1 \otimes Z^{x_1})$ component of the input state is multiplied by $\lambda_{x_1,y_1}^{Q_1}$ and transformed into the $Q_1 \otimes Z^{y_1}$ component after applying the first MCM. This component is then twisted by H_1 into $\mathcal{G}^\dagger(Q_2 \otimes Z^{x_2})$ and so on... In the end the component becomes $Q_l \otimes Z^{y_l}$, which is then estimated by the terminating measurement. In terms of Pauli weights, this sequence corresponds exactly to the transversal of the nodes $v_0, \ldots, v_l$ along the specified path, with the Pauli fidelities (edges) acquired along the way. Those unshown components of the evolving underlying density matrix do not affect the result, as they are either undetected, eliminated by randomized compiling, or averaged out during data processing.

estimate

$$\log\left(\frac{\lambda_M^{v_l}}{\lambda_M^{v_0}}\prod_{i=1}^l \lambda_{x_i,y_i}^{Q_i}\right) \equiv -\log\lambda_M^{v_0} + \sum_{i=1}^l \log\lambda_{x_i,y_i}^{Q_i} + \log\lambda_M^{v_l}. \tag{25}$$

For the modification, simply replace step 5 by the following:

(5*) Prepare ρ and perform compiled terminating measurement for observable $\mathcal{G}^\dagger(Q_1 \otimes Z^{x_1})$. Record the result as r'. Repeat multiple times and estimate the expectation value $t = \mathbb{E}[r']$.

(6*) Output $\log s/t$.

We note that the x_i, y_i, and Q_is are the input to the algorithm. H_is can then be calculated via Eq. (23). The circuits are then determined and executed nonadaptively, independent of the measurement outcomes observed. The final output is calculated based on the measurement outcomes and the inputs.

In the following, we consider the modified version of the protocol. First we prove that the output is indeed the desired quantity.

Theorem 1 (Main result). For any path $v_0, e_{x_1,y_1}^{Q_1}, v_1, \ldots,$ $e_{x_l,y_l}^{Q_l}, v_l$, the function $-v_0 + \sum_{i=1}^l e_{x_i,y_i}^{Q_i} + v_l$ is learnable.

We note that our protocol outputs the path with its two end points (SPAM errors). When the provided path forms a directed cycle (i.e., $v_0 = v_l$), one can immediately see that the output $\sum_{i=1}^l \log\lambda_{x_i,y_i}^{Q_i}$ does not contain SPAM error parameters since they have been canceled. In such

cases, one may also concatenate the directed cycle with itself L times and run the protocol on the extended path $v_0, e_{x_1,y_1}^{Q_1}, v_1, \ldots, e_{x_l,y_l}^{Q_l}, v_0, e_{x_1,y_1}^{Q_1}, v_1, \ldots, e_{x_l,y_l}^{Q_l}, \ldots, v_0$. (In fact, as mentioned in Sec. IV A, this is the usual method used in CB-type algorithms.) Then the protocol will output $L\sum_{i=1}^l \log\lambda_{x_i,y_i}^{Q_i}$, from which one can retrieve the product of fidelities. One can further perform the experiment for varying L and run a regression to get the original product of fidelities.

Proof of Theorem 1. It suffices to prove the correctness of our protocol. The probability of observing MCM results $m_1, \ldots, m_l$ is the trace of resulting unnormalized density matrix

$$p_{m_1,\ldots,m_l} = \langle\!\langle I | \mathcal{T}_{m_l} H_{l-1} \cdots H_1 \mathcal{T}_{m_1} | \rho \rangle\!\rangle. \tag{26}$$

Here I means the identity matrix. Conditioned on observing this result, the quantum state is

$$|\sigma_{m_1,\ldots,m_l}\rangle\!\rangle = \frac{1}{p_{m_1,\ldots,m_l}}\mathcal{T}_{m_l} H_{l-1} \cdots H_1 \mathcal{T}_{m_1} | \rho \rangle\!\rangle. \tag{27}$$

Hence by Lemma 7, the expectation of r conditioned on observing the measurement results is

$$\mathbb{E}[r|m_1,\ldots,m_l] = \lambda_M^{v_l}\langle\!\langle Q_l \otimes Z^{y_l} | \sigma_{m_1,\ldots,m_l}\rangle\!\rangle. \tag{28}$$

Thus

$$\mathbb{E}\left[(-1)^{\sum_{i=1}^{l} m_i \cdot (x_i + y_i)} r\right] \tag{29}$$

$$= \sum_{m_1,\dots,m_l \in \mathbb{Z}_2^n} p_{m_1,\dots,m_l} (-1)^{\sum_{i=1}^{l} m_i \cdot (x_i + y_i)} \mathbb{E}\left[r | m_1, \dots, m_l\right] \tag{30}$$

$$= \sum_{m_1,\dots,m_l \in \mathbb{Z}_2^n} \lambda_M^{v_l} (-1)^{\sum_{i=1}^{l} m_i \cdot (x_i + y_i)} \langle\!\langle Q_l \otimes Z^{y_l} | \mathcal{T}_{m_l} H_{l-1} \cdots H_1 \mathcal{T}_{m_1} | \rho \rangle\!\rangle \tag{31}$$

$$= \frac{\lambda_M^{v_l}}{2^{(2n+m)l}} \sum_{\substack{m_1,\dots,m_l \in \mathbb{Z}_2^n, a_1,\dots,a_l \in \mathbb{Z}_2^n \\ b_1,\dots,b_l \in \mathbb{Z}_2^n, P_1,\dots,P_l \in \mathcal{P}^m}} (-1)^{\sum_{i=1}^{l} m_i \cdot (x_i + y_i + a_i + b_i)} \prod_{i=1}^{l} \lambda_{a_i,b_i}^{P_i}$$

$$\times \langle\!\langle Q_l \otimes Z^{y_l} | P_l \otimes Z^{b_l} \rangle\!\rangle \langle\!\langle \mathcal{G}^\dagger (P_l \otimes Z^{a_l}) | \mathcal{H}_{l-1}(P_{l-1} \otimes Z^{b_{l-1}}) \rangle\!\rangle \cdots \langle\!\langle \mathcal{G}^\dagger (P_1 \otimes Z^{a_1}) | \rho \rangle\!\rangle \tag{32}$$

$$= \lambda_M^{v_l} \sum_{\substack{a_1,\dots,a_l \in \mathbb{Z}_2^n \\ P_1,\dots,P_l \in \mathcal{P}^m}} \prod_{i=1}^{l} \lambda_{a_i,x_i+y_i+a_i}^{P_i} f(a_1,\dots,a_l, P_1,\dots,P_l) \langle\!\langle \mathcal{G}^\dagger (P_1 \otimes Z^{a_1}) | \rho \rangle\!\rangle, \tag{33}$$

where function f takes value $0, \pm 1$ depending on the inner products

$$2^{(n+m)l} f(a_1,\dots,a_l, P_1,\dots,P_l)$$
$$= \langle\!\langle Q_l \otimes Z^{y_l} | P_l \otimes Z^{x_l + y_l + a_l} \rangle\!\rangle \langle\!\langle \mathcal{G}^\dagger (P_l \otimes Z^{a_l}) | \mathcal{H}_{l-1}(P_{l-1} \otimes Z^{x_{l-1}+y_{l-1}+a_{l-1}}) \rangle\!\rangle$$
$$\cdots \langle\!\langle \mathcal{G}^\dagger (P_2 \otimes Z^{a_2}) | \mathcal{H}_1(P_1 \otimes Z^{x_1+y_1+a_1}) \rangle\!\rangle. \tag{34}$$

Equation (32) uses Lemma 1.

Next, we prove that only one term is left in the summation, that is, $f(a_1,\dots,a_l, P_1,\dots,P_l) = \prod_{i=1}^{l} \delta_{a_i,x_i} \delta_{P_i,Q_i}$. Here, δ is the Kronecker δ function that takes value 1 only when its two subscripts are equal, and takes value 0 otherwise. We prove this inductively. If f is non-zero, then from the first inner product we have $P_l = Q_l$, $a_l = x_l$. Since $\mathcal{H}_{l-1}(Q_{l-1} \otimes Z^{y_{l-1}}) = \mathcal{G}^\dagger(Q_l \otimes Z^{x_l})$, from the second inner product we have $P_{l-1} = Q_{l-1}$, $a_{l-1} = x_{l-1} \dots$ In the end we have $P_1 = Q_1$, $a_1 = x_1$. Hence f is nonzero only when $P_i = Q_i$ and $a_i = x_i$, and it takes value 1 in this case. Putting this result back we get

$$\mathbb{E}\left[(-1)^{\sum_{i=1}^{l} m_i \cdot (x_i + y_i)} r\right] = \lambda_M^{v_l} \langle\!\langle \mathcal{G}^\dagger (Q_1 \otimes Z^{x_1}) | \rho \rangle\!\rangle \prod_{i=1}^{l} \lambda_{x_i,y_i}^{Q_i}. \tag{35}$$

This is the expected value of s. On the other hand, by Lemma 7 the expected value of t is $\lambda_M^{v_0} \langle\!\langle \mathcal{G}^\dagger (Q_1 \otimes Z^{x_1}) | \rho \rangle\!\rangle$, hence $\log s/t$ is an estimate of the desired quantity. ∎

V. LEARNABILITY OF MCMs

We have seen in Theorem 1 that certain combinations of Pauli fidelities can be learned using our generalized CB algorithm. Now we show that, under our noise model these turn out to be all information that can be SPAM-robustly learned about noisy MCMs *via any algorithm*. For this purpose, we develop a theory on the learnability of MCMs, generalizing the framework from Ref. [15] about the learnability of noisy Clifford gates.

To start with, let us formally summarize our noise model assumptions:

(1) All single-qubit gates can be noiselessly implemented.

(2) A set of multiqubit Clifford gates $\{\mathcal{G}_i\}$ can be implemented followed by (unknown) gate-dependent Pauli noise channels $\{\Lambda_{\mathcal{G}_i}\}$.

(3) A noisy MCM $\{\mathcal{T}_k\}$ with (unknown) Pauli fidelities $\{\lambda_{x,y}^Q\}$ can be applied.

(4) An unknown but fixed initial state ρ can be prepared [28].

(5) Any POVM can be measured following an unknown (symmetric) Pauli noise channel Λ_M. See Appendix A 2 for details.

(6) All Pauli fidelities of noise channels are strictly positive. All noise channels are not at the boundary of the completely positive polytope.

Assumption 1 is standard for randomized-compiling-based protocols [5,6,21], and can be relaxed such that all layers of single-qubit gates have *gate-independent noise*. Assumptions 2–4 can be enforced using randomized compiling. Note that, if we exclude assumption 3, the noise model reduces to the standard Pauli noise model with Clifford gates, for which the learnability has been studied in Ref. [15]. We also remark that, it is straightforward to generalize our learnability theory to allow multiple distinct noisy MCM gadgets. We omit that for conciseness. Assumption 6 is mostly for mathematical convenience: the first part basically says that the noise is not overwhelmingly large; the second part ensures that all noise parameters can be perturbed without making the noise model nonphysical.

To learn the noise parameters in the noise model $\mathcal{N} = \{\rho, \{\Lambda_{\mathcal{G}_i}\}, \{\lambda_{x,y}^Q\}, \Lambda_M\}$, the most general form of experiments one can perform is to prepare the initial state, apply a sequence of gates and MCM gadgets, and perform a terminating measurement. Any experiment maps a (realization of the) noise model to a probability distribution over the measurement outcomes (from both the MCMs and the terminating measurements). We say two noise models $\mathcal{N}_1, \mathcal{N}_2$ are *indistinguishable* if for every possible experiment, they yield the same probability distribution. Otherwise they are *distinguishable*.

We are interested in which parameters of the noise model are learnable from experiments. Formally, a function f on noise models maps a noise model $\mathcal{N}$ to a real number, denoted as $f(\mathcal{N})$. For example, a 0 chain can be viewed as a function of noise models that reflects terminating measurement noises $\left(\sum \epsilon_i v_i\right)(\mathcal{N}) = \sum \epsilon_i \log \lambda_M^{v_i}$. Similarly, a 1 chain can be viewed as a function of noise models that reflects MCM noises $\left(\sum \epsilon_{x,y}^Q e_{x,y}^Q\right)(\mathcal{N}) = \sum \epsilon_{x,y}^Q \log \lambda_{x,y}^Q$.

A function f is called *learnable* [15] if

$$\forall \mathcal{N}_1, \mathcal{N}_2 : \mathcal{N}_1 \text{ is indistinguishable from } \mathcal{N}_2$$
$$\Rightarrow f(\mathcal{N}_1) = f(\mathcal{N}_2).$$

Otherwise, f is *unlearnable*. This definition is as expected, because the ability to learn an unlearnable function would imply the ability to distinguish indistinguishable noise models, which leads to a contradiction. From this definition alone, f being learnable is only a necessary condition for the existence of an experiment to actually learn its value. Here we will prove that our learning algorithm can indeed learn any of such learnable functions within arbitrary precision. We also remark that a function being unlearnable is a fundamental limitation for *any learning protocols*, not just specific to CB-type protocols.

A. Learnability of Pauli fidelities

The protocol presented in Sec. IV C is important in helping us understand the learnability of Pauli fidelities. To make a complete characterization of the learnable information, first we need the following lemma.

Lemma 3 (Lemma 1, Supplemental Material of Ref. [15]). Denote the set of all learnable 1 chains by F_L. Then F_L forms a linear subspace of the edge space C.

The lemma states that learnability defined above is closed under linear operations. For completeness we present a proof here.

Proof. Given any $\mu_1, \mu_2 \in F_L, \forall \mathcal{N}_1, \mathcal{N}_2, \forall \alpha \neq 0$,

$$\mathcal{N}_1 \text{ is indistinguishable from } \mathcal{N}_2$$
$$\Rightarrow \mu_1(\mathcal{N}_1) = \mu_1(\mathcal{N}_2) \text{ and } \mu_2(\mathcal{N}_1) = \mu_2(\mathcal{N}_2)$$
$$\Rightarrow (\mu_1 + \alpha\mu_2)(\mathcal{N}_1) = (\mu_1 + \alpha\mu_2)(\mathcal{N}_2). \quad (36)$$

Thus $\mu_1 + \alpha\mu_2 \in F_L$. Note that the last line uses the linearity of 1 chains. ∎

Corollary 1. For any 1 chain μ, $\mu + \partial\mu$ is learnable.

Proof. As a special case of Theorem 1, for any edge e, $e + \partial e$ is learnable. Hence by linearity, $\mu + \partial\mu$ is learnable. ∎

Corollary 2. 1 chains in the cycle space Z are learnable.

Proof. For 1 chains $\mu \in Z$, by Corollary 1, $\mu + \partial\mu$ is learnable. By definition of cycles, $\partial\mu = 0$, thus μ is learnable. ∎

We remark that our focus is on MCMs (1 chains), and we are not interested in terminating measurement errors (0 chains). Thus the implication of Corollary 2 is that cycle space can be learned unaffected by state-preparation and terminating measurement noises, that is, SPAM robustly.

Furthermore, we will show that the cycle space is all the information that can be learned SPAM robustly, and thus our protocol learns all the information that can be learned.

Theorem 2. The protocol in Theorem 1 is complete in the sense that the space of learnable information F_L is equal to the cycle space Z.

Proof. $Z \subseteq F_L$ is already proved in Corollary 2, so it remains to show that $F_L \subseteq Z$. By Lemma 2, it suffices to show that F_L is orthogonal to the cut space U. We will show that every cut vector induces a gauge transformation that convert one noise model to another indistinguishable noise model. Since learnable functions should be invariant

under such transformation by definition, we can conclude that they should be orthogonal to the cut space.

Recall that a generic experiment begins with an initial state ρ, followed by a sequence of MCMs interleaved by single-qubit or multiqubit Clifford gates $\mathcal{C}_i$ (if we extend our noise model to include non-Clifford multiqubit gates, they can appear in the circuit, too), and finally a terminating measurement. Assume that there are d layers of MCMs with outcome denoted by $m_1, \ldots, m_d$, and that the outcome of the terminating measurement is denoted by m_*, the probability distribution of the outcomes is given by

$$\Pr[m_1, \ldots, m_d, m_*]$$
$$= \text{Tr}(E_{m_*}(\mathcal{C}_d \circ \mathcal{T}_{m_d} \circ \cdots \circ \mathcal{C}_1 \circ \mathcal{T}_{m_1} \circ \mathcal{C}_0(\rho))). \quad (37)$$

Here, each $\mathcal{C}_i$ can be a layer of noiseless single-qubit gates, or a noisy multiqubit Clifford gates, or a concatenation of both. E_{m_*} is the POVM element of the noisy terminating measurement.

Thus we can see that, for any invertible linear map $\mathcal{D}$, the following gauge transformation does not change the probability distribution.

$$\begin{cases} \rho \mapsto \mathcal{D}(\rho) \\ \mathcal{C}_i \mapsto \mathcal{D}\mathcal{C}_i\mathcal{D}^{-1} \\ \mathcal{T}_k \mapsto \mathcal{D}\mathcal{T}_k\mathcal{D}^{-1} \\ E_j \mapsto (\mathcal{D}^{-1})^\dagger(E_j) \end{cases}. \quad (38)$$

For a noise model $\mathcal{N}_1$, we can use this to construct an indistinguishable model $\mathcal{N}_2$ from $\mathcal{N}_1$.

Given any cut vector $\delta(\nu) \in U$, since $\delta(\sum \nu) = 0$, without loss of generality assume $\nu = \sum \epsilon_{v_i} v_i$ has coefficient 0 on vertex $\bigcirc_{m+n}$. Let $\mathcal{D}$ be the Pauli diagonal map defined by

$$\forall P \in \mathcal{P}^{n+m}, \quad \mathcal{D}(P) = \eta^{\epsilon_{\text{pt}(P)}} P \quad (39)$$

for a positive $\eta \neq 1$. We hence need to verify that $\mathcal{N}_2$ still satisfies all of our model assumptions:

(1) If $\mathcal{C}$ is a layer of single-qubit gate, it preserves the pattern of any input Pauli operator [29], thus $\mathcal{C}_i = \mathcal{D}\mathcal{C}_i\mathcal{D}^{-1}$, i.e., single-qubit gates remains noiseless.

(2) If $\mathcal{C}$ is a multiqubit Clifford gate $\mathcal{G}_i$ followed by a Pauli noise channel $\Lambda_{\mathcal{G}_i}$, we have

$$\mathcal{D}\Lambda_{\mathcal{G}_i}\mathcal{G}_i\mathcal{D}^{-1} = \left(\mathcal{D}\Lambda_{\mathcal{G}_i}\mathcal{G}_i\mathcal{D}^{-1}\mathcal{G}_i^\dagger\right)\mathcal{G}_i = \Lambda'_{\mathcal{G}_i}\mathcal{G}_i, \quad (40)$$

which is still the same Clifford gate followed by a Pauli channel.

(3) From Lemma 1, we can easily see that $\{\mathcal{D}\mathcal{T}_k\mathcal{D}^{-1}\}$ remains a compiled MCM (i.e., $\{\mathcal{D}\mathcal{U}_k\mathcal{D}^{-1}\}$ remains a uniform stochastic instrument), with Pauli fidelities changed according to

$$\log\lambda^Q_{x,y} \mapsto \log\lambda^Q_{x,y} + (\epsilon_{\text{pt}(Q\otimes Z^y)} - \epsilon_{\text{pt}(\mathcal{G}^\dagger(Q\otimes Z^x))})\log\eta$$
$$= \log\lambda^Q_{x,y} + \langle\delta(\nu), e^Q_{x,y}\rangle\log\eta \quad (41)$$

Note that since $e^I_{0,0}$ is always a self-loop on vertex 0, the trace-preserving condition $\log\lambda^I_{0,0} = 0$ is preserved.

(4) The initial state transforms from ρ to $\mathcal{D}(\rho)$. For the terminating measurement, if the ideal POVM is $\{F_j\}$, its noisy implementation is $\{\Lambda^\dagger_M(F_j)\}$ where Λ_M is a symmetric Pauli channel. The transformation results in $\Lambda_M \mapsto \Lambda_M\mathcal{D}^{-1}$, which is still a symmetric Pauli channel.

(5) Finally, it is not hard to see Assumption 5 still holds under the transformation, as long as η is sufficiently close to 1.

Thus $\mathcal{N}_2$ satisfies all of our assumptions. Moreover, for any 1 chain μ, thanks to Eq. (41) we have

$$\mu(\mathcal{N}_2) = \mu(\mathcal{N}_1) + \langle\delta(\nu), \mu\rangle\log\eta. \quad (42)$$

Hence if $\mu \in F_L$, by the definition of learnable functions, we must have $\langle\delta(\nu), \mu\rangle = 0$. This completes the proof. ■

B. Learnability of Pauli error rates

Theorem 2 gives a complete classification for the learnable Pauli fidelities. Since Pauli error rates have a clearer physical meaning, naturally we would ask for a classification for the learnable Pauli error rates. In this subsection, we give a partial classification under low noise rate assumption. Under low noise rate assumption, the Pauli fidelities are close to 1, thus we can make approximation

$$p^P_{a,b} = \frac{1}{4^{n+m}}\sum_{x,y\in\mathbb{Z}_2^n, Q\in\mathcal{P}^m}(-1)^{a\cdot x+b\cdot y+\langle P,Q\rangle}\lambda^Q_{x,y} \quad (43)$$

$$\approx \frac{1}{4^{n+m}}\sum_{x,y\in\mathbb{Z}_2^n, Q\in\mathcal{P}^m}(-1)^{a\cdot x+b\cdot y+\langle P,Q\rangle}(1 + \log\lambda^Q_{x,y}) \quad (44)$$

$$= \frac{1}{4^{n+m}}\sum_{x,y\in\mathbb{Z}_2^n, Q\in\mathcal{P}^m}(-1)^{a\cdot x+b\cdot y+\langle P,Q\rangle}\log\lambda^Q_{x,y} + \delta_{a,0}\delta_{b,0}\delta_{P,I}. \quad (45)$$

Equation (44) uses the fact that when $x \approx 1$, $x \approx 1 + \log x$, and the δ in Eq. (45) means the Kronecker δ function.

As an example, the learnable Pauli error rates for $n = m = 1$ MCMs with $G = $ CNOT are

$$p_{1,1}^I, p_{1,1}^Z, p_{1,1}^X, p_{1,1}^Y$$
$$p_{0,0}^I + p_{0,0}^Z, p_{0,1}^I + p_{1,0}^Z + p_{0,1}^Z + p_{1,0}^I$$
$$p_{0,0}^X - p_{0,0}^Y, p_{1,0}^X - p_{1,0}^Y, p_{0,1}^X - p_{0,1}^Y, p_{0,1}^X - p_{0,1}^Z$$
$$p_{0,1}^I + p_{1,0}^I - p_{0,0}^I, p_{0,1}^X + p_{1,0}^X + p_{0,0}^X, p_{0,0}^I + p_{1,0}^I + p_{0,1}^X.$$

More generally, we have the following propositions. Proofs are deferred to Sec. C.

Proposition 1. $\forall G, P, \forall a \neq 0, b \neq 0, p_{a,b}^P$ is learnable.

The proof is existential, but Corollary 1 also suggests a concrete way to learn this quantity. Specifically, one can learn each $\log \lambda_{x,y}^Q$ as in Eq. (45) together with its end point using our protocol. Thanks to Corollary 1, the end points will cancel as we sum up all the estimators. Alternatively, one can try to directly decompose Eq. (45) into sum of cycles. As a concrete example, we give the protocol for learning $p_{1,1}^I$ when $G = $ CNOT (the setting for Fig. 4). $p_{1,1}^I$ can be decomposed as

$$16p_{1,1}^I \approx \log \lambda_{0,0}^I + (\log \lambda_{1,1}^X + \log \lambda_{1,1}^Y + \log \lambda_{0,0}^Y + \log \lambda_{0,1}^Z + \log \lambda_{0,0}^X$$
$$+ \log \lambda_{0,0}^Z + \log \lambda_{0,1}^Z + \log \lambda_{1,1}^I + \log \lambda_{1,1}^Z) - (\log \lambda_{0,1}^Y + \log \lambda_{1,0}^Y + \log \lambda_{0,1}^Z$$
$$+ \log \lambda_{0,1}^X + \log \lambda_{1,0}^Y + \log \lambda_{0,1}^Z + \log \lambda_{1,0}^I + \log \lambda_{0,1}^I + \log \lambda_{1,0}^Z + \log \lambda_{0,1}^Z). \tag{46}$$

We note that the first term, $\log \lambda_{0,0}^I$ is fixed to be 0, so we only need to apply our protocol 2 times to learn the second and the last term.

An important application for MCMs is syndrome measurement. In this case we measure the stabilizer S by introducing an ancilla, perform a Hadamard gate on the ancilla, apply controlled S using the ancilla as control, and then apply another Hadamard transformation on the ancilla and measure the ancilla. See Fig. 6 for an illustration of the ideal Clifford gate G in the case when we want to measure stabilizers $S_1, \ldots, S_n$ simultaneously. Similar noise extraction under syndrome measurements settings have been studied in Refs. [30,31], though under different noise assumptions.

The following two propositions consider this case. For $k \in \mathbb{Z}_2^n$, define $S^k = \prod S_i^{k_i}$. Denote the set of Pauli operators that commute with all stabilizers by $C_{\mathcal{P}^m}(\mathcal{S})$.

The summation over $k \in \mathbb{Z}_2^n$ can be intuitively understood as the following. Since we are measuring stabilizers, it makes no effect if we randomly apply some stabilizers before the measurement. Without loss of generality suppose we indeed applied the random stabilizers. Then errors that differ by stabilizers become equivalent, thus only their average error rates can be learnable.

We note that when $S_1 = \cdots = S_n = I$, or equivalently, $G = I$ and we are implementing a subsystem measurement, this proposition shows the learnability of $p_{0,0}^I$, which is the process fidelity between the uniform stochastic instrument and the ideal subsystem measurement [32].

As another special case, when $n = 0$, no stabilizers are measured. By the requirements on G, G must be identity. Now $C_{\mathcal{P}^m}(\mathcal{S}) = \mathcal{P}^m$, and this proposition degenerates into the fact that for an isolated Pauli channel, every parameter is learnable.

Proposition 2. $\forall P \in C_{\mathcal{P}^m}(\mathcal{S}), \sum_{k \in \mathbb{Z}_2^n} p_{0,0}^{S^k P}$ is learnable.

Proposition 3. $\forall P \in C_{\mathcal{P}^m}(\mathcal{S}), \forall a \neq 0, \sum_{k \in \mathbb{Z}_2^n} p_{0,a}^{S^k P} + p_{a,0}^{S^k P}$ is learnable.
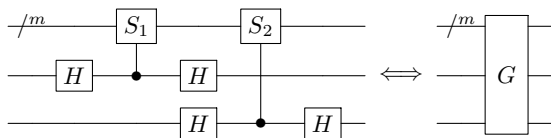

FIG. 6. Illustration of the ideal Clifford gate G for syndrome measurements. Here we take $n = 2$.

The summation over $k \in \mathbb{Z}_2^n$ originates from the random stabilizer intuition. The intuitive meaning of summing $p_{0,a}^P$ and $p_{a,0}^P$ together is that we cannot distinguish between wrong but consistent with postmeasurement state result and a correct but inconsistent with postmeasurement state result. These two possibilities can only be learned as a unity.

VI. APPLICATION: TESTING THE INDEPENDENCE OF MEASUREMENT AND STATE PREPARATION

In Sec. III B we described the measure and prepare instrument model. The key assumption we made is that the measurement and the subsequent state preparation are independent and suffers from uncorrelated noises. In experiment, we can test whether the assumption is true or not by checking the additional structures of the model. To see why we can do this, we need more understanding on the structure.

Lemma 4. $\exists \zeta, \xi$ s.t. $\forall x, y \in \mathbb{Z}_2^n, Q \in \mathcal{P}^m, \lambda_{x,y}'^Q = \zeta_x^Q \xi_y^Q$ if and only if $\forall x_1, x_2, y_1, y_2 \in \mathbb{Z}_2^n, Q \in \mathcal{P}^m$, the correlation $c_{x_1,x_2,y_1,y_2}^Q := \log \lambda_{x_1,y_1}'^Q + \log \lambda_{x_2,y_2}'^Q - \log \lambda_{x_2,y_1}'^Q - \log \lambda_{x_1,y_2}'^Q = 0$.

Proof. The only if part is easy. Suppose $\exists \zeta, \xi$ s.t. $\forall x, y \in \mathbb{Z}_2^n, Q \in \mathcal{P}^m, \lambda_{x,y}'^Q = \zeta_x^Q \xi_y^Q$, then for $\forall x_1, x_2, y_1, y_2 \in \mathbb{Z}_2^n, Q \in \mathcal{P}^m$,

$$\log \lambda_{x_1,y_1}'^Q + \log \lambda_{x_2,y_2}'^Q - \log \lambda_{x_2,y_1}'^Q - \log \lambda_{x_1,y_2}'^Q \quad (47)$$

$$= \log \zeta_{x_1}^Q + \log \xi_{y_1}^Q + \log \zeta_{x_2}^Q + \log \xi_{y_2}^Q - \log \zeta_{x_2}^Q$$
$$- \log \xi_{y_1}^Q - \log \zeta_{x_1}^Q - \log \xi_{y_2}^Q = 0. \quad (48)$$

Conversely, for the if part, set $x_1 = y_1 = 0, x_2 = x, y_2 = y$ we have

$$\log \lambda_{x,y}'^Q = \log \lambda_{x,0}'^Q + \log \lambda_{0,y}'^Q - \log \lambda_{0,0}'^Q. \quad (49)$$

Thus we can set $\zeta_x^Q = \lambda_{x,0}'^Q / \sqrt{\lambda_{0,0}'^Q}, \xi_y^Q = \lambda_{0,y}'^Q / \sqrt{\lambda_{0,0}'^Q}$, proving the result. ∎

From Lemma 4 we see that to test the independence of measurement and state preparation, we can equivalently test whether the correlations are all zero. Furthermore, these conditions are learnable information and hence are indeed verifiable.

Lemma 5. $\forall x, y \in \mathbb{Z}_2^n, Q \in \mathcal{P}^m, \log \lambda_{x_1,y_1}'^Q + \log \lambda_{x_2,y_2}'^Q - \log \lambda_{x_2,y_1}'^Q - \log \lambda_{x_1,y_2}'^Q$ is learnable.

Proof.

$$\partial \left(e_{x_1,y_1}^Q + e_{x_2,y_2}^Q - e_{x_2,y_1}^Q - e_{x_1,y_2}^Q \right)$$

$$= -\text{pt}(\mathcal{G}^\dagger(Q \otimes Z_1^x)) + \text{pt}(Q \otimes Z_1^y)$$
$$- \text{pt}(\mathcal{G}^\dagger(Q \otimes Z_2^x)) + \text{pt}(Q \otimes Z_2^y)$$
$$- \text{pt}(\mathcal{G}^\dagger(Q \otimes Z_2^x)) + \text{pt}(Q \otimes Z_1^y)$$
$$- \text{pt}(\mathcal{G}^\dagger(Q \otimes Z_1^x)) + \text{pt}(Q \otimes Z_2^y) = 0. \quad (50)$$

∎

Since $c_{x_1,x_2,y_1,y_2}^Q = -c_{x_2,x_1,y_1,y_2}^Q = -c_{x_1,x_2,y_2,y_1}^Q$, in experiments we need only to estimate for $x_1 < x_2$ and $y_1 < y_2$ (comparison under alphabetical order).

VII. NUMERICAL SIMULATIONS

In this section, we perform numerical simulations for the simple example of $G = \text{CNOT}$ and $m = n = 1$, which is a basic gadget for syndrome measurement.

A. Learning Pauli fidelities

We randomly generate a noisy quantum instrument and simulate our (modified) protocol on it. For each of the cycle basis, the way we learn it is to concatenate the cycle multiple times in the main circuit (steps 1–4) so that the cycle becomes of length $\ell = 12$ (i.e., 12 MCMs in each of the main circuit). We assign values for the random variables in the randomized compilings 100 times. This results in 100 random circuits, which we call compiled circuits, and for each compiled circuit we sample 100 shots. We take the same number of shots for the auxiliary circuit (step 5), so in total $N = 20\,000$ shots are taken for each data point. We then process the data to get the geometric mean of the Pauli fidelities in the cycle basis. The standard deviations of the obtained results are calculated through bootstrapping. Our results are shown in Fig. 7.

In the setting above, the variance can also be estimated theoretically. Suppose s, t, λ are all sufficiently close to 1, the variance is roughly given by [33]

$$\text{Var}\left[e^{\frac{1}{\ell} \log \frac{s}{t}}\right] \approx \text{Var}\left[\frac{s-t}{\ell}\right] = \frac{\text{Var}(s) + \text{Var}(t)}{\ell^2}$$

$$= \frac{1 - \mathbb{E}[s]^2 + 1 - \mathbb{E}[t]^2}{\ell^2 N/2}$$

$$= \frac{1 - \lambda_0^2 \lambda^{2\ell} + 1 - \lambda_0^2}{\ell^2 N/2} \approx \frac{2(1 - \lambda_0^2)}{\ell^2 N/2}$$

$$+ \frac{\lambda_0^2 (1 - \lambda^2)}{\ell N/2}. \quad (51)$$

Here $\ell = lL$ is the number of MCMs used in each shot, N is the total number of shots used for estimating s and t, and $\lambda_0 = \langle\!\langle \mathcal{G}^\dagger(Q_1 \otimes Z^{x_1}) | \rho \rangle\!\rangle \lambda_M^{vl}$ is a constant related only to SPAM errors. We have used the assumption that the noise rate is small, so s and t concentrates around values close to 1.

B. Testing independence of measurement and state preparation

We then simulate the experiment of testing the independence of the measurement and state preparation, as described in Sec. VI. We randomly construct a quantum instrument for a noisy MCM and apply randomized compiling on it. A noisy MCM with independent Pauli channel
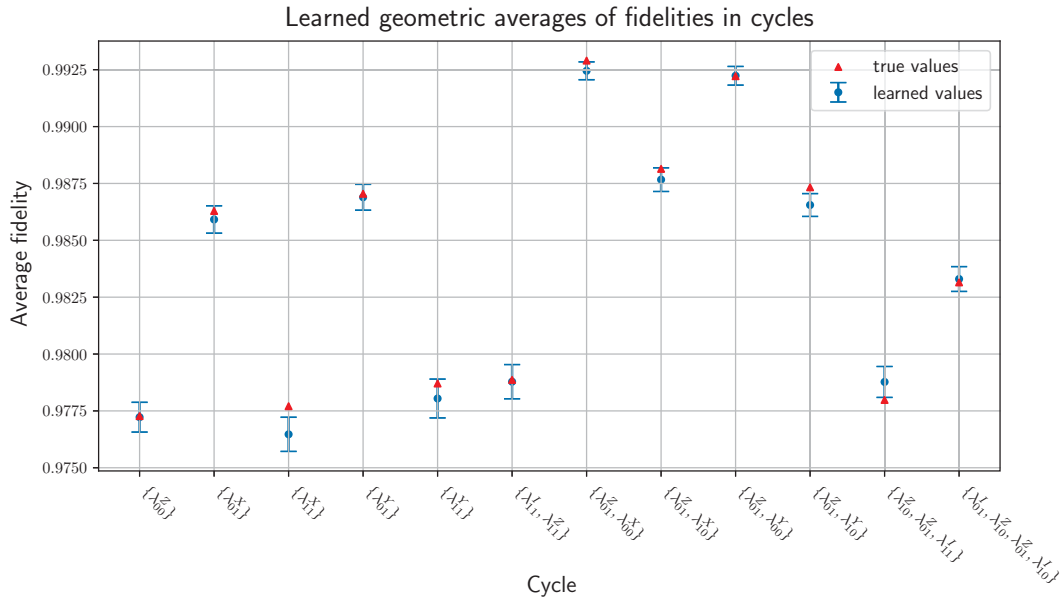
FIG. 7.   Simulation results in Sec. VII A. There are 13 learnable degrees of freedom, each corresponds to a cycle. The x axis shows the corresponding cycles while the y axis shows the learned geometric average of the fidelities in the cycle. The learnable information λ_{00}^{I} is omitted here since it is always 1. The error bars represent one standard deviation.

noise before and after measurement (a measure and prepare instrument) is used for comparison. The two MCMs have roughly the same noise level and we simulate our test for them. Specifically, we learn the correlations $c_{0,1,0,1}^{Q}$ for $Q \in \{I, X, Y, Z\}$. Because of the antisymmetry of the correlations, these are all the correlations that we need to test. We use an approach slightly different to the previous part to learn the correlations. For each correlation, we learn the 4 log fidelities separately using the (modified) protocol and add or subtract them together. By Lemma 5 we know that

the boundaries will cancel out. For compiled MCM, 100 compiled circuits are used for the main circuit, each with 20 000 shots. 2 000 000 shots are used for the auxiliary circuit, so one log fidelity is learned using 4 000 000 shots. The same number of shots are used for measure and prepare instrument. The standard deviation is estimated via bootstrapping. Our results are shown in Fig. 8. One can see that our criteria judges the independence nicely. We note that the true values are always non-negative because of the positivity requirement of the quantum instruments.
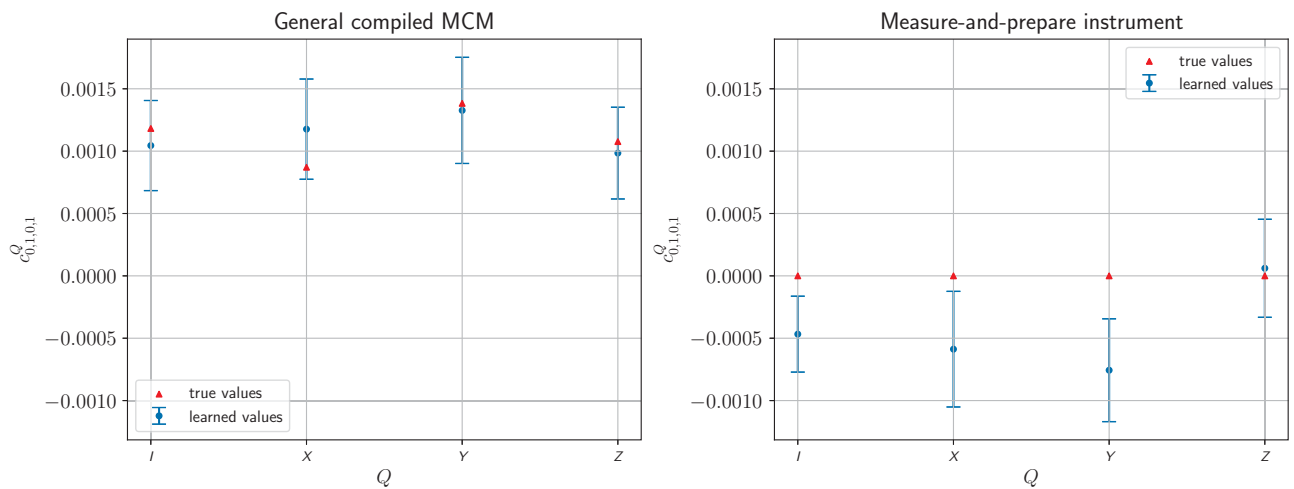

FIG. 8.   Simulation results in VII B. The left side shows the estimated correlations of a general compiled MCM while the right side shows that of a measure and prepare channel. The y axis represents the learned values of the correlation $c_{0,1,0,1}^{Q}$. The x axis indicates the corresponding Q. The error bars represent one standard deviation. Logarithms are in base e.

C. Learning Pauli error rate

Finally we learn the $p_{1,1}^I$ of a random quantum instrument. The method has already been used as an example in Eq. (46). We learn the two terms in Eq. (46) by the (modified) protocol, each with 100 compiled circuits for the main circuit and we take ten shots from each of the compiled circuits. 1000 shots are taken for the auxiliary circuit. We repeat the above experiment 10 times and use their mean as output, so in total 40 000 shots are taken. The standard deviation is estimated from repetitions. In our simulation, the true value for $p_{1,1}^I$ is 2.79×10^{-4}, while our learned value is 2.60×10^{-4} with estimated standard deviation 0.55×10^{-4}.

VIII. DISCUSSION AND OUTLOOK

In this work, we conduct a comprehensive investigation on learning noisy MCMs. Using graph-theoretic tools, we can determine all the learnable degrees of freedom for a compiled MCM, and design a CB-type protocol to learn all learnable information. We demonstrate our learning protocol in a numerical example. As an application, we illustrate how our protocol can be used to test the independence between measurement and state preparation for an MCM.

Our results put forward a fundamental limit on characterizing randomly compiled MCMs. Namely, we show that certain parameters are coupled with gauge transformations, thus cannot be precisely learned SPAM robustly. In practice, it could be desirable to obtain knowledge about those "unlearnable" degrees of freedom. For example, one might want to know whether the dominant error is in the measurement readout or in the postmeasurement state, which might provide insight on how to improve the hardware design. Solving such a problem would require additional assumptions to anchor the gauge, e.g., assuming noiseless state preparation or terminating measurements, or use the physicality constraints as investigated in Ref. [15]. How to characterize MCMs efficiently under such physical assumptions requires dedicated exploration in the future.

Moreover, the concatenation technique discussed at the end of Sec. IV C may not always be applicable. In Sec. V A we proved that all the learnable information are cycles. For CB with interleaved gates, we can prove that the connected components of the pattern-transfer graph are strongly connected. As a consequence, it permits a directed cycle basis, meaning that all the learnable information can be decomposed into directed cycles and thus concatenated. However, we cannot obtain similar results for our protocol. Whether the entire cycle space can be learned through concatenation (i.e., whether the pattern transfer graph admits a circuit basis) remains an open question.

Finally, although we have given a complete characterization of the learnability of noisy MCMs in terms of Pauli fidelities, we are only able to give a partial characterization for the learnability of Pauli error rates with three special cases in Sec. V B. For noisy Clifford gate, in contrast, the learnability of Pauli error rates and Pauli fidelities are basically the same, thanks to the fact that the cycle space is invariant under Walsh-Hadamard transform (see Sec. D for details). Such a property breaks down when we include MCMs. Therefore, a more comprehensive and physically motivating understanding of the learnability for Pauli error rates is left for future investigation.

Note added. Recently, an independent work proposed a similar algorithm [34].

The code for numerical simulation can be found at https://github.com/zhihan-z/MCM_Learnability.

APPENDIX A: RANDOMIZED COMPILING

1. Randomized compiling for MCMs

In this section, we explain how randomized compiling techniques can reduce a general quantum instrument $\mathcal{M} = \{\mathcal{M}_k\}$ into an instrument $\mathcal{T} = \{\mathcal{T}_k\}$ that fits our model. The scheme we adopt is a simple variation of the scheme proposed in Ref. [21]. Figure 9 illustrates our randomized compiling scheme.

Specifically, we choose $P \in \mathcal{P}^m, \alpha, \beta, \gamma \in \mathbb{Z}_2^n$ uniformly randomly and define operator $P_1 \otimes P_2 = \mathcal{G}^\dagger(P \otimes Z^\beta X^\alpha)$. We apply gate $P_1 \otimes P_2$, followed by the noisy MCM $\mathcal{M}$, and then gate $P \otimes X^\alpha Z^\gamma$. The classical output from the MCM is added by α to form the measurement result of the compiled channel. We make the assumption that all single-qubit unitaries can be implemented noiselessly. This is justified by the fact that noises of multiqubit
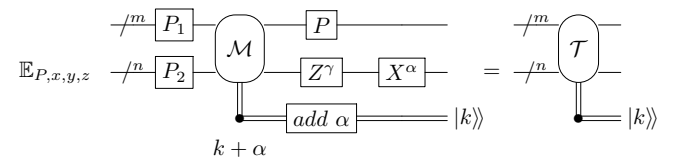

FIG. 9. Illustration of the randomized compiling scheme.

operations are usually much higher than that of single-qubit operations. Then, after the compiling, the effective instrument becomes

$$
\mathcal{T}_k = \mathop{\mathbb{E}}_{P,\alpha,\beta,\gamma} \left[(\mathcal{P} \otimes \mathcal{X}^\alpha \mathcal{Z}^\gamma) \mathcal{M}_{k-\alpha} (\mathcal{G}^\dagger (\mathcal{P} \otimes \mathcal{Z}^\beta \mathcal{X}^\alpha) \mathcal{G}) \right]
$$
(A1)

$$
= \mathop{\mathbb{E}}_{P,\alpha,\beta,\gamma} \left[(\mathcal{P} \otimes \mathcal{X}^\alpha \mathcal{Z}^\gamma) \mathcal{M}_{k-\alpha} \mathcal{G}^\dagger (\mathcal{P} \otimes \mathcal{Z}^\beta \mathcal{X}^\alpha) \right] \mathcal{G}.
$$
(A2)

Here, gates and channels are used interchangeably. To simplify, we need a lemma from Ref. [21].

Lemma 6 (Theorem 2, Ref. [21]). The following randomized compiling twirls a general quantum instrument $\mathcal{M} = \{\mathcal{M}_k\}$ to a uniform stochastic instrument $\{\mathcal{U}_k\}$:

$$
\mathop{\mathbb{E}}_{P,\alpha,\beta,\gamma} \left[(\mathcal{P} \otimes \mathcal{X}^\alpha \mathcal{Z}^\gamma) \mathcal{M}_{k-\alpha} \mathcal{G}^\dagger (\mathcal{P} \otimes \mathcal{Z}^\beta \mathcal{X}^\alpha) \right] = \mathcal{U}_k.
$$
(A3)

Intuitively, the random α averages over different absolute measurement outcomes, making the $\Lambda_{a,b}$ in $\mathcal{U}_k$ [Eq. (9)] independent of k and depends instead only on the relative differences a and b.

From Lemma 6, we immediately see that $\mathcal{T}_k = \mathcal{U}_k \mathcal{G}$ and thus justifies our model. We note that this compiling scheme is slightly different from Fig. 5 of Ref. [21] as we are not allowed to insert twirling gates between the Clifford gate G and the subsystem measurement since we assume that they are implemented as a whole.

2. Randomized compiling for terminating measurements

We model a terminating measurement as a (special) Pauli channel Λ_M, followed by an ideal terminating measurement. The Pauli channel here has the form

$$
\Lambda_M = \sum_{F \in \mathcal{P}^{n+m}} \lambda_M^{\mathrm{pt}(F)} |F\rangle\!\rangle\langle\!\langle F|.
$$
(A4)

One can immediately see that this Pauli channel is special in the sense that the Pauli fidelities are indexed by Pauli weight patterns instead of Pauli operators. We note that this is for mathematical convenience, and our proof also works for general Pauli channels.

Again, our noise model can be realized via randomized compiling. Assume the noisy terminating measurements implemented can be viewed as an arbitrary channel $\mathcal{E}^M$ followed by an ideal terminating measurement. Then we can use the standard randomized compiling [5] techniques to twirl the noise into a special Pauli channel Λ^M. More specifically, suppose we want to measure observable E, then we shall select a random $H \in \mathcal{C}^{\otimes n+m}$, apply gate H and then measure in $\mathcal{H}(E)$ basis.

Lemma 7. If we use the above compiled terminating measurements to measure observable $E \in \mathcal{P}^{n+m}$ on ρ, then the expectation of the result is $\lambda_M^{\mathrm{pt}(E)} \langle\!\langle E|\rho\rangle\!\rangle$. Here λ_M are actually Pauli fidelities of the terminating measurement noise. It is indexed by Pauli weight patterns.

Proof. The expectation value is

$$
\mathop{\mathbb{E}}_{H \in \mathcal{C}^{\otimes n+m}} [\mathrm{Tr}(\mathcal{H}(E) \cdot \mathcal{E} \circ \mathcal{H}(\rho))]
$$

$$
= \mathrm{Tr} \left(E \mathop{\mathbb{E}}_{H \in \mathcal{C}^{\otimes n+m}} \left[\mathcal{H}^\dagger \circ \mathcal{E} \circ \mathcal{H}(\rho) \right] \right) = \lambda_M^{\mathrm{pt}(E)} \langle\!\langle E|\rho\rangle\!\rangle.
$$
(A5)

The second equality is the direct consequence of Schur's lemma [35] as in standard randomized compiling. ∎

APPENDIX B: STRUCTURE OF THE MEASURE-AND-PREPARE INSTRUMENTS

In this section, we formalize the definition of measure-and-prepare instrument and express it in the dual space. The process is shown in Fig. 1(c) and follows the description in Sec. I B.

We start with a uniform stochastic instrument [we use the dual-space form Eq. (20) here]. Without the Clifford gate G, a uniform stochastic instrument is

$$
\mathcal{U}_k = \frac{1}{2^{2n+m}} \sum_{x,y \in \mathbb{Z}_2^n, Q \in \mathcal{P}^m} (-1)^{k \cdot (x+y)} \lambda_{x,y}^Q |Q \otimes Z^y\rangle\!\rangle\langle\!\langle Q \otimes Z^x|.
$$
(B1)

Suppose we obtain the measurement outcome k, which corresponds to $\mathcal{U}_k$ being applied. A measure-and-prepare instrument corresponds to tracing out the ancilla register, followed by a preparation of fresh ancilla with X correction [Fig. 1(c)]. Our goal is to derive the dual-space form of this process.

First, tracing out the ancilla register in $\mathcal{U}_k$, we get

$$
\frac{1}{2^{n+m}} \sum_{x \in \mathbb{Z}_2^n, Q \in \mathcal{P}^m} (-1)^{k \cdot x} \lambda_{x,0}^Q |Q\rangle\!\rangle\langle\!\langle Q \otimes Z^x|
$$

$$
= \frac{1}{2^{n+m}} \sum_{x \in \mathbb{Z}_2^n, Q \in \mathcal{P}^m} (-1)^{k \cdot x} \zeta_x^Q |Q\rangle\!\rangle\langle\!\langle Q \otimes Z^x|.
$$
(B2)

Here we have defined $\zeta_x^Q = \lambda_{x,0}^Q$. Next, we model the state-preparation procedure as preparing a perfect initial state $|0^n\rangle\!\rangle$ (the computational basis state where all qubits are zero) followed by a Pauli noise channel, so the next step is to put the refreshed ancillas in,

$$\frac{1}{2^{2n+m}} \sum_{x,y\in\mathbb{Z}_2^n,Q\in\mathcal{P}^m} (-1)^{k\cdot x}\zeta_x^Q |Q\otimes Z^y\rangle\!\rangle\langle\!\langle Q\otimes Z^x|. \tag{B3}$$

For the Pauli-noise channel, denote its Pauli fidelity for $Q\otimes Z^y$ by ξ_y^Q. Applying this noise, we get

$$\frac{1}{2^{2n+m}} \sum_{x,y\in\mathbb{Z}_2^n,Q\in\mathcal{P}^m} (-1)^{k\cdot x}\zeta_x^Q\xi_y^Q |Q\otimes Z^y\rangle\!\rangle\langle\!\langle Q\otimes Z^x|. \tag{B4}$$

Apply the NOT gates X^k, we get the effective channel for measurement result k

$$\mathcal{U}_k' = \frac{1}{2^{2n+m}} \sum_{x,y\in\mathbb{Z}_2^n,Q\in\mathcal{P}^m} (-1)^{k\cdot(x+y)}\zeta_x^Q\xi_y^Q |Q\otimes Z^y\rangle\!\rangle\langle\!\langle Q\otimes Z^x|. \tag{B5}$$

$\{\mathcal{U}_k'\}$ is the measure-and-prepare instrument expressed in dual space. Compared with Eq. (13), one can see that $\{\mathcal{U}_k'\}$ is also a uniform stochastic instrument, with the additional structure

$$\lambda_{x,y}'^Q = \zeta_x^Q\xi_y^Q, \tag{B6}$$

i.e., the Pauli fidelity is factorized into a product.

In terms of error rates, define q and r as the inverse Fourier transformation of ζ and ξ over $\mathbb{Z}_2^n$, respectively,

$$q_a^P = \frac{1}{2^{n+2m}} \sum_{x\in\mathbb{Z}_2^n,Q\in\mathcal{P}^m} (-1)^{a\cdot x+\langle P,Q\rangle}\zeta_x^Q, \quad r_b^P = \frac{1}{2^{n+2m}} \sum_{x\in\mathbb{Z}_2^n,Q\in\mathcal{P}^m} (-1)^{b\cdot x+\langle P,Q\rangle}\xi_x^Q. \tag{B7}$$

q and r corresponds to the Pauli error rates for the measurement-and-prepare process, respectively. The inverse transformations are

$$\zeta_x^Q = \sum_{a\in\mathbb{Z}_2^n,P\in\mathcal{P}^m} (-1)^{a\cdot x+\langle P,Q\rangle}q_a^P, \quad \xi_x^Q = \sum_{a\in\mathbb{Z}_2^n,P\in\mathcal{P}^m} (-1)^{b\cdot x+\langle P,Q\rangle}r_b^P. \tag{B8}$$

And we have

$$p_{a,b}'^P = \frac{1}{4^{n+m}} \sum_{x,y\in\mathbb{Z}_2^n,Q\in\mathcal{P}^m} (-1)^{a\cdot x+b\cdot y+\langle P,Q\rangle}\lambda_{x,y}'^Q \tag{B9}$$

$$= \frac{1}{4^{n+m}} \sum_{x,y\in\mathbb{Z}_2^n,Q\in\mathcal{P}^m} (-1)^{a\cdot x+b\cdot y+\langle P,Q\rangle}\zeta_x^Q\xi_y^Q \tag{B10}$$

$$= \frac{1}{4^{n+2m}} \sum_{x,y\in\mathbb{Z}_2^n,Q_1,Q_2,P\in\mathcal{P}^m} (-1)^{a\cdot x+b\cdot y+\langle P,Q_1Q_2\rangle+\langle P,Q_2\rangle}\zeta_x^{Q_1}\xi_y^{Q_2} \tag{B11}$$

$$= \frac{1}{4^{n+2m}} \sum_{x,y\in\mathbb{Z}_2^n,Q_1,Q_2,P\in\mathcal{P}^m} (-1)^{a\cdot x+b\cdot y+\langle P,Q_1\rangle+\langle PP_1,Q_2\rangle}\zeta_x^{Q_1}\xi_y^{Q_2} \tag{B12}$$

$$= \sum_{P_1\cdot P_2=P} \left(\frac{1}{2^{n+2m}} \sum_{x\in\mathbb{Z}_2^n,Q\in\mathcal{P}^m} (-1)^{a\cdot x+\langle P,Q\rangle}\zeta_x^Q\right)\left(\frac{1}{2^{n+2m}} \sum_{x\in\mathbb{Z}_2^n,Q\in\mathcal{P}^m} (-1)^{b\cdot x+\langle P,Q\rangle}\xi^Q\right) \tag{B13}$$

$$= \sum_{P_1\cdot P_2=P} q_a^{P_1}r_b^{P_2}. \tag{B14}$$

Equation (B11) uses the fact that $\sum_{P\in\mathcal{P}^m}(-1)^{\langle P,Q_1Q_2\rangle}$ is nonzero only when $Q_1=Q_2$ and takes value 4^m in this case. By plugging in this equation to Eq. (8) we get the physical space expression of the measure-and-prepare instrument.

$$\mathcal{U}'_k = \sum_{a,b\in\mathbb{Z}_2^n, P\in\mathcal{P}^m, P_1\cdot P_2=P} q_a^{P_1} r_b^{P_2}(\mathcal{P}\otimes|k+b\rangle\!\rangle\langle\!\langle k+a|). \tag{B15}$$

APPENDIX C: PROOFS FROM SEC. V B

Proof of Proposition 1. It suffices to show the corresponding 1 chain is in Z.

$$\partial\frac{1}{4^{n+m}}\sum_{x,y\in\mathbb{Z}_2^n, Q\in\mathcal{P}^m}(-1)^{a\cdot x+b\cdot y+\langle P,Q\rangle}e_{x,y}^Q \tag{C1}$$

$$=\frac{1}{4^{n+m}}\sum_{x,y\in\mathbb{Z}_2^n, Q\in\mathcal{P}^m}(-1)^{a\cdot x+b\cdot y+\langle P,Q\rangle}(-\mathrm{pt}(\mathcal{G}^\dagger(Q\otimes Z^x))+\mathrm{pt}(Q\otimes Z^y)) \tag{C2}$$

$$=-\frac{1}{4^{n+m}}\sum_{x\in\mathbb{Z}_2^n, Q\in\mathcal{P}^m}(-1)^{a\cdot x+\langle P,Q\rangle}\mathrm{pt}(\mathcal{G}^\dagger(Q\otimes Z^x))\sum_{y\in\mathbb{Z}_2^n}(-1)^{b\cdot y}$$

$$+\frac{1}{4^{n+m}}\sum_{y\in\mathbb{Z}_2^n, Q\in\mathcal{P}^m}(-1)^{b\cdot y+\langle P,Q\rangle}\mathrm{pt}(Q\otimes Z^y)\sum_{x\in\mathbb{Z}_2^n}(-1)^{a\cdot x}=0. \tag{C3}$$

Note that the addition of Pauli weight patterns are performed in the sense of 0 chains. Equation (C2) uses the definition of boundary operator and Eq. (C3) uses the fact that when $a\neq 0$ and $b\neq 0$, $\sum_{y\in\mathbb{Z}_2^n}(-1)^{b\cdot y}=\sum_{x\in\mathbb{Z}_2^n}(-1)^{a\cdot x}=0$. ∎

Proof of Proposition 2.

$$\partial\frac{1}{4^{n+m}}\sum_{x,y,k\in\mathbb{Z}_2^n, Q\in\mathcal{P}^m}(-1)^{\langle S^k P,Q\rangle}e_{x,y}^Q \tag{C4}$$

$$=\frac{1}{4^{n+m}}\sum_{x,y\in\mathbb{Z}_2^n, Q\in\mathcal{P}^m}(-1)^{\langle P,Q\rangle}\partial e_{x,y}^Q\sum_{k\in\mathbb{Z}_2^n}(-1)^{\langle S^k,Q\rangle} \tag{C5}$$

$$=\frac{2^n}{4^{n+m}}\sum_{x,y\in\mathbb{Z}_2^n, Q\in C_{\mathcal{P}^m}(\mathcal{S})}(-1)^{\langle P,Q\rangle}(-\mathrm{pt}(\mathcal{G}^\dagger(Q\otimes Z^x))+\mathrm{pt}(Q\otimes Z^y)) \tag{C6}$$

$$=\frac{2^n}{4^{n+m}}\sum_{x,y\in\mathbb{Z}_2^n, Q\in C_{\mathcal{P}^m}(\mathcal{S})}(-1)^{\langle P,Q\rangle}(-\mathrm{pt}(S^x Q\otimes Z^x)+\mathrm{pt}(Q\otimes Z^y)) \tag{C7}$$

$$=-\frac{1}{4^m}\sum_{x\in\mathbb{Z}_2^n, Q\in C_{\mathcal{P}^m}(\mathcal{S})}(-1)^{\langle P,Q\rangle}\mathrm{pt}(S^x Q\otimes Z^x)+\frac{1}{4^m}\sum_{y\in\mathbb{Z}_2^n, Q\in C_{\mathcal{P}^m}(\mathcal{S})}(-1)^{\langle P,Q\rangle}\mathrm{pt}(Q\otimes Z^y) \tag{C8}$$

$$=-\frac{1}{4^m}\sum_{x\in\mathbb{Z}_2^n, Q\in C_{\mathcal{P}^m}(\mathcal{S})}(-1)^{\langle P,S^x Q\rangle}\mathrm{pt}(Q\otimes Z^x)+\frac{1}{4^m}\sum_{x\in\mathbb{Z}_2^n, Q\in C_{\mathcal{P}^m}(\mathcal{S})}(-1)^{\langle P,Q\rangle}\mathrm{pt}(Q\otimes Z^x)=0. \tag{C9}$$

Equation (C6) uses the fact that $\sum_{k\in\mathbb{Z}_2^n}(-1)^{\langle S^k,Q\rangle}$ is nonzero and takes value 2^n only when $Q\in C_{\mathcal{P}^m}(\mathcal{S})$ and Eq. (C9) makes a substitution $Q\mapsto S^x Q$ for the first term and also uses the fact $P\in C_{\mathcal{P}^m}(\mathcal{S})$. ∎

Proof of Proposition 3.

$$\partial \frac{1}{4^{n+m}} \sum_{Q\in\mathcal{P}^m,x,y,k\in\mathbb{Z}_2^n} (-1)^{\langle S^kP,Q\rangle} \left[(-1)^{a\cdot x} + (-1)^{a\cdot y}\right] e_{x,y}^Q \tag{C10}$$

$$= \frac{1}{4^{n+m}} \sum_{Q\in\mathcal{P}^m,x,y\in\mathbb{Z}_2^n} (-1)^{\langle P,Q\rangle} \left[(-1)^{a\cdot x} + (-1)^{a\cdot y}\right] \partial e_{x,y}^Q \sum_{k\in\mathbb{Z}_2^n} (-1)^{\langle S^k,Q\rangle} \tag{C11}$$

$$= \frac{2^n}{4^{n+m}} \sum_{Q\in C_{\mathcal{P}^m}(\mathcal{S}),x,y\in\mathbb{Z}_2^n} (-1)^{\langle P,Q\rangle} \left[(-1)^{a\cdot x} + (-1)^{a\cdot y}\right] \partial e_{x,y}^Q \tag{C12}$$

$$= -\frac{2^n}{4^{n+m}} \sum_{Q\in C_{\mathcal{P}^m}(\mathcal{S}),x,y\in\mathbb{Z}_2^n} (-1)^{\langle P,Q\rangle} \left[(-1)^{a\cdot x} + (-1)^{a\cdot y}\right] \mathrm{pt}(S^x Q \otimes Z^x)$$

$$+ \frac{2^n}{4^{n+m}} \sum_{Q\in C_{\mathcal{P}^m}(\mathcal{S}),x,y\in\mathbb{Z}_2^n} (-1)^{\langle P,Q\rangle} \left[(-1)^{a\cdot x} + (-1)^{a\cdot y}\right] \mathrm{pt}(Q \otimes Z^y) \tag{C13}$$

$$= -\frac{1}{4^m} \sum_{Q\in C_{\mathcal{P}^m}(\mathcal{S}),x\in\mathbb{Z}_2^n} (-1)^{a\cdot x+\langle P,Q\rangle} \mathrm{pt}(S^x Q \otimes Z^x) + \frac{1}{4^m} \sum_{Q\in C_{\mathcal{P}^m}(\mathcal{S}),y\in\mathbb{Z}_2^n} (-1)^{a\cdot y+\langle P,Q\rangle} \mathrm{pt}(Q \otimes Z^y) \tag{C14}$$

$$= -\frac{1}{4^m} \sum_{Q\in C_{\mathcal{P}^m}(\mathcal{S}),x\in\mathbb{Z}_2^n} (-1)^{a\cdot x+\langle P,S^x Q\rangle} \mathrm{pt}(Q \otimes Z^x) + \frac{1}{4^m} \sum_{Q\in C_{\mathcal{P}^m}(\mathcal{S}),x\in\mathbb{Z}_2^n} (-1)^{a\cdot x+\langle P,Q\rangle} \mathrm{pt}(Q \otimes Z^x) = 0. \tag{C15}$$

Equation (C12) uses the trick similar to Eq. (C6) and Eq. (C15) uses the trick similar to Eq. (C9). ∎

APPENDIX D: INVARIANCE OF CYCLE SPACE UNDER WALSH-HADAMARD TRANSFORMATION

In Ref. [15], the authors made a conjecture about the learnability of Pauli error rates for noisy Clifford gates. They conjectured that in this case, the cycle space is invariant under Walsh-Hadamard transformation. That is, for a learnable cycle, if we substitute the log fidelities by the corresponding Pauli error rates, the resulting quantity about Pauli error rates is also learnable (under first-order approximation). Though noisy Clifford gates is not the focus of this paper and from Sec. V B obviously this special property no longer holds in our generalized case, to complete the physical vs dual-space picture, in this section we give a proof for this conjecture.

We note by the way that though in a different language, Lemma 3 in Ref. [12] effectively gives a constructive proof for the theorem for the special case where no interleaving single-qubit Clifford gates are allowed. But in this section,

we follow the nonconstructive proof fashion in Sec. V B to prove the general theorem.

Once again, we emphasize that in this section we follow our convention and model the noise as happening after the Clifford gate, so our definition of pattern-transfer graph is different from that of Ref. [15]. Reference [15] has proved that for Pauli fidelities, the cycle space is learnable while the cut space is unlearnable, and their results are still valid using our convention.

Theorem 3. For a cycle $v_0, e^{Q_1}, v_1, \ldots, e^{Q_l}, v_0$, $\sum_{k=1}^{l} p^{Q_k}$ is learnable.

Proof. Since we are considering a cycle, view the subscript 0 equivalent as the subscript l. By definition of pattern-transfer graph, $\exists H_1, \ldots, H_l \in \mathcal{C}^{\otimes m}$ ($\mathcal{C}$ is the group of single-qubit Clifford gates) such that $H_{k-1}Q_{k-1}H_{k-1}^\dagger = G^\dagger Q_k G$. Then it suffices to prove the learnability of the corresponding 1 chain.

$$\partial \frac{1}{4^m} \sum_{k=1}^{l} \sum_{R\in\mathcal{P}^m} (-1)^{\langle Q_k,R\rangle} e^R \tag{D1}$$

$$= \frac{1}{4^m} \sum_{R \in \mathcal{P}^m} \partial e^R \sum_{k=1}^{l} (-1)^{\langle Q_k, R \rangle} \tag{D2}$$

$$= \frac{1}{4^m} \sum_{R \in \mathcal{P}^m} \text{pt}(R) \sum_{k=1}^{l} (-1)^{\langle Q_k, R \rangle} - \frac{1}{4^m} \sum_{R \in \mathcal{P}^m} \text{pt}(\mathcal{G}^\dagger(R)) \sum_{k=1}^{l} (-1)^{\langle Q_k, R \rangle} \tag{D3}$$

$$= \frac{1}{4^m} \sum_{R \in \mathcal{P}^m} \text{pt}(R) \sum_{k=1}^{l} (-1)^{\langle Q_k, R \rangle} - \frac{1}{4^m} \sum_{R \in \mathcal{P}^m} \text{pt}(R) \sum_{k=1}^{l} (-1)^{\langle GH_{k-1}Q_{k-1}H_{k-1}^\dagger G^\dagger, GRG^\dagger \rangle} \tag{D4}$$

$$= \frac{1}{4^m} \sum_{R \in \mathcal{P}^m} \text{pt}(R) \sum_{k=1}^{l} (-1)^{\langle Q_k, R \rangle} - \frac{1}{4^m} \sum_{R \in \mathcal{P}^m} \text{pt}(R) \sum_{k=1}^{l} (-1)^{\langle Q_{k-1}, H_{k-1}^\dagger R H_{k-1} \rangle} \tag{D5}$$

$$= \frac{1}{4^m} \sum_{R \in \mathcal{P}^m} \text{pt}(R) \sum_{k=1}^{l} (-1)^{\langle Q_k, R \rangle} - \frac{1}{4^m} \sum_{R \in \mathcal{P}^m} \text{pt}(R) \sum_{k=1}^{l} (-1)^{\langle Q_{k-1}, R \rangle} \tag{D6}$$

$$= 0. \tag{D7}$$

Equation (D3) uses the definition of pattern-transfer graph for cycle benchmarking. Equation (D4) makes the substitution $R \mapsto GRG^\dagger$. Equation (D5) uses the fact that conjugating by the same Clifford gate does not change the commutation relation. Equation (D6) makes the substitution $R \mapsto H_{k-1}RH_{k-1}^\dagger$ and uses the fact that single-qubit Clifford gates do not change patterns. ∎

By simple counting of degrees of freedom, [15] proved that Theorem 3 implies that the learnability of Pauli error rates are also completely characterized by cycle space. That is, the cycle space is the space of all learnable about Pauli error rates.

---

[1] C. Ryan-Anderson, J. G. Bohnet, K. Lee, D. Gresh, A. Hankin, J. Gaebler, D. Francois, A. Chernoguzov, D. Lucchetti, N. C. Brown, *et al.*, Realization of real-time fault-tolerant quantum error correction, Phys. Rev. X **11**, 041058 (2021).

[2] K. Singh, C. E. Bradley, S. Anand, V. Ramesh, R. White, and H. Bernien, Mid-circuit correction of correlated phase errors using an array of spectator qubits, Science **380**, 1265 (2023).

[3] I. Aleiner, R. Allen, T. I. Andersen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, J. Atalaya, R. Babbush, D. Bacon, *et al.*, Suppressing quantum errors by scaling a surface code logical qubit, Nature **614**, 676 (2023).

[4] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, *et al.*, Logical quantum processor based on reconfigurable atom arrays, Nature **626**, 58 (2024).

[5] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, Phys. Rev. A **94**, 052325 (2016).

[6] A. Hashim, R. K. Naik, A. Morvan, J.-L. Ville, B. Mitchell, J. M. Kreikebaum, M. Davis, E. Smith, C. Iancu, K. P. O'Brien, *et al.*, Randomized compiling for scalable quantum computing on a noisy superconducting quantum processor, Phys. Rev. X **11**, 041039 (2021).

[7] S. T. Flammia and J. J. Wallman, Efficient estimation of Pauli channels, ACM Trans. Quantum Comput. **1**, 1 (2020).

[8] A. Erhard, J. J. Wallman, L. Postler, M. Meth, R. Stricker, E. A. Martinez, P. Schindler, T. Monz, J. Emerson, and R. Blatt, Characterizing large-scale quantum computers via cycle benchmarking, Nat. Commun. **10**, 5347 (2019).

[9] J. Emerson, M. Silva, O. Moussa, C. Ryan, M. Laforest, J. Baugh, D. G. Cory, and R. Laflamme, Symmetrized characterization of noisy quantum processes, Science **317**, 1893 (2007).

[10] S. T. Flammia, in *17th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2022)*, edited by F. Le Gall and T. Morimae, Leibniz International Proceedings in Informatics (LIPIcs) (Schloss Dagstuhl—Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2022), Vol. 232, p. 4:1.

[11] E. Van Den Berg, Z. K. Minev, A. Kandala, and K. Temme, Probabilistic error cancellation with sparse Pauli–Lindblad models on noisy quantum processors, Nat. Phys. **19**, 1116 (2023).

[12] A. Carignan-Dugas, D. Dahlen, I. Hincks, E. Ospadov, S. J. Beale, S. Ferracin, J. Skanes-Norman, J. Emerson, and J. J. Wallman, The error reconstruction and compiled calibration of quantum computing cycles, ArXiv:2303.17714.

[13] L. Govia, P. Jurcevic, C. Wood, N. Kanazawa, S. Merkel, and D. McKay, A randomized benchmarking suite for midcircuit measurements, New J. Phys. **25**, 123016 (2023).

[14] K. Rudinger, G. J. Ribeill, L. C. Govia, M. Ware, E. Nielsen, K. Young, T. A. Ohki, R. Blume-Kohout, and T. Proctor, Characterizing midcircuit measurements on a superconducting qubit using gate set tomography, Phys. Rev. Appl. **17**, 014014 (2022).

[15] S. Chen, Y. Liu, M. Otten, A. Seif, B. Fefferman, and L. Jiang, The learnability of Pauli noise, Nat. Commun. **14**, 52 (2023).

[16] Note that, while SPAM typically refers to "state preparation and measurements," in our context, we treat MCMs as a midcircuit gadget different from the usual destructive measurement. Thus, the "M" in SPAM should be interpreted specifically as the terminating measurements.

[17] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Gate set tomography, Quantum **5**, 557 (2021).

[18] E. Bäumer, V. Tripathi, D. S. Wang, P. Rall, E. H. Chen, S. Majumder, A. Seif, and Z. K. Minev, Efficient long-range entanglement using dynamic circuits, PRX Quantum **5**, 030339 (2024).

[19] A. Hashim, A. Carignan-Dugas, L. Chen, C. Juenger, N. Fruitwala, Y. Xu, G. Huang, J. Wallman, and I. Siddiqi, Quasi-probabilistic readout correction of mid-circuit measurements for adaptive feedback via measurement randomized compiling, ArXiv:2312.14139.

[20] M. McEwen, D. Kafri, Z. Chen, J. Atalaya, K. Satzinger, C. Quintana, P. V. Klimov, D. Sank, C. Gidney, A. Fowler, *et al.*, Removing leakage-induced correlated errors in superconducting quantum error correction, Nat. Commun. **12**, 1761 (2021).

[21] S. J. Beale and J. J. Wallman, Randomized compiling for subsystem measurements, ArXiv:2304.06599.

[22] This is the symplectic inner product between the symplectic representation of P and Q. Specifically, the symplectic representation ϕ of a Pauli operator $P \in \mathcal{P}^m$ is the unique length $2m$ bit string that satisfies $P = i^{\phi(P)^T \Upsilon \phi(P)} X^{\phi(P)_{\text{odd}}} \cdot Z^{\phi(P)_{\text{even}}}$, where $\phi(P)_{\text{odd}}$ and $\phi(P)_{\text{even}}$ denotes the substring of $\phi(P)$ consists of odd and even positions, respectively, and $\Upsilon = \bigoplus_{j=1}^m \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$. The inner product is defined as the binary symplectic form $\langle P, Q \rangle = \phi(P)^T (\Upsilon + \Upsilon^T) \phi(Q)$ mod 2.

[23] D. Greenbaum, Introduction to quantum gate set tomography, ArXiv:1509.02921.

[24] M. M. Wilde, From classical to quantum Shannon theory, ArXiv:1106.1445.

[25] Note that, our labeling convention is slightly different from Ref. [15] because in our model, the noise

happens after the Clifford gate, while in Ref. [15] the noise happens before the Clifford gate. As a result, the pattern-transfer graph defined in Ref. [15] is (V, E) with $V = \mathbb{Z}_2^m$ and $E = \{e^Q := (\text{pt}(Q), \text{pt}(\mathcal{G}(Q)) | Q \in \mathcal{P}^m\}$, while our pattern-transfer graph is defined as (V, E) with $V = \mathbb{Z}_2^{n+m}$ and $E = \{e^Q_{x,y} := (\text{pt}(\mathcal{G}^\dagger(Q \otimes Z^x)), \text{pt}(Q \otimes Z^y)) | x, y \in \mathbb{Z}_2^n, Q \in \mathcal{P}^m\}$.

[26] F. Harary, *Graph Theory (on Demand Printing Of 02787)* (CRC Press, Boca Raton, 2018).

[27] B. Bollobás, *Modern Graph Theory* (Springer Science & Business Media, New York, 1998), Vol. 184.

[28] We note that regarding the initial state, though it does not affect our learnability theory, for our protocol the overlap $\langle\!\langle \mathcal{G}^\dagger(Q_1 \otimes Z^{x_1}) | \rho \rangle\!\rangle$ do affect the sample complexity. It is desirable to choose and prepare an initial state with high overlap $\langle\!\langle \mathcal{G}^\dagger(Q_1 \otimes Z^{x_1}) | \rho \rangle\!\rangle$.

[29] More precisely, since $\mathcal{C}$ is allowed to be non-Clifford, it can map an input Pauli to multiple output Paulis, each of which has the same pattern as the input.

[30] T. Wagner, H. Kampermann, D. Bruß, and M. Kliesch, Pauli channels can be estimated from syndrome measurements in quantum error correction, Quantum **6**, 809 (2022).

[31] T. Wagner, H. Kampermann, D. Bruß, and M. Kliesch, Learning logical Pauli noise in quantum error correction, Phys. Rev. Lett. **130**, 200601 (2023).

[32] D. McLaren, M. A. Graydon, and J. J. Wallman, Stochastic errors in quantum instruments, ArXiv:2306.07418.

[33] It is worth noting that a rigorous formula for the variance is unattainable as our protocol has some small probability of failure. In extreme cases where either s or t in the protocol deviates significantly from their expected value such that s/t becomes negative, the output of the protocol is undefined. However, assuming that the number of shots taken is sufficient so that such extreme cases can be ignored and that the noise rate is low, the variance of the results can be estimated.

[34] J. Hines and T. Proctor, Pauli noise learning for mid-circuit measurements, ArXiv:2406.09299.

[35] W. Fulton and J. Harris, *Representation Theory: A First Course* (Springer Science & Business Media, New York, 2013), Vol. 129.